

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 7, 2002, 10:33:38 : Search time 20.04 Seconds  
(without alignments)  
1784.385 Million cell updates/sec

Title: US-09-842-256-2

Perfect score: 7631  
Sequence: 1 MSGMGENTSPSRARTRRK.....NQLPGMDIKQEGDTTRKCYC 1464

Scoring table:

BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/2/1aa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/2/1aa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/2/1aa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/2/1aa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/2/1aa/PTUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/2/1aa/Backfilltest.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7631	100.0	1464	US-08-891-640-2	Sequence 2, Appli
2	1037.5	13.6	1036	US-08-891-640-3	Sequence 3, Appli
3	317	4.2	846	US-08-885-291-55	Sequence 55, Appli
4	317	4.2	846	US-09-107-847-2	Sequence 55, Appli
5	317	4.2	846	US-09-496-672-55	Sequence 55, Appli
6	309.5	4.1	855	US-08-816-693A-2	Sequence 2, Appli
7	309.5	4.1	855	US-08-885-291-2	Sequence 2, Appli
8	309.5	4.1	855	US-09-496-672-2	Sequence 2, Appli
9	292	3.8	870	US-08-785-241-4	Sequence 2, Appli
10	289	3.8	816	US-08-785-310A-8	Sequence 4, Appli
11	288	3.8	816	US-08-816-693A-53	Sequence 53, Appli
12	288	3.8	816	US-08-885-291-53	Sequence 53, Appli
13	288	3.8	816	US-09-496-672-53	Sequence 53, Appli
14	286.5	3.7	875	US-08-971-188-8	Sequence 8, Appli
15	283.5	3.6	875	US-08-785-241-5	Sequence 8, Appli
16	277	3.6	2414	US-08-227-536-2	Sequence 2, Appli
17	277	3.6	2414	PCT-US95-04682-2	Sequence 2, Appli
18	269.5	3.5	826	US-08-785-241-6	Sequence 2, Appli
19	269.5	3.5	826	US-08-480-473B-2	Sequence 6, Appli
20	269.5	3.5	826	US-08-915-213-2	Sequence 2, Appli
21	269.5	3.5	826	US-09-148-547-2	Sequence 2, Appli
22	269.5	3.5	826	US-09-235-217-2	Sequence 2, Appli
23	269.5	3.5	826	PCT-US96-10251-2	Sequence 2, Appli
24	269	3.5	626	US-08-971-188-10	Sequence 10, Appli
25	267	3.5	747	US-08-816-693A-51	Sequence 51, Appli
26	267	3.5	747	US-08-885-291-51	Sequence 51, Appli
27	267	3.5	747	US-09-496-672-51	Sequence 51, Appli

28	265	3.5	824	3	US-08-885-291-52	Sequence 52, Appli
29	265	3.5	824	4	US-09-496-672-52	Sequence 52, Appli
30	262	3.4	824	2	US-08-816-693A-52	Sequence 52, Appli
31	259	3.4	824	2	US-08-785-310A-7	Sequence 7, Appli
32	258.5	3.4	3969	4	US-08-061-376-5	Sequence 5, Appli
33	252	3.3	2441	1	US-08-194-468-2	Sequence 2, Appli
34	252	3.3	2441	3	US-08-961-739-2	Sequence 2, Appli
35	248	3.2	805	2	US-08-480-473B-4	Sequence 4, Appli
36	248	3.2	805	3	US-08-915-213-4	Sequence 4, Appli
37	248	3.2	805	4	US-09-235-217-4	Sequence 4, Appli
38	248	3.2	805	5	PCT-US96-10251-4	Sequence 4, Appli
39	245.5	3.2	810	1	US-08-785-241-7	Sequence 7, Appli
40	235	3.1	848	1	US-08-045-806-4	Sequence 4, Appli
41	235	3.1	848	1	US-08-366-051B-4	Sequence 4, Appli
42	234.5	3.1	1187	1	US-08-320-559-28	Sequence 28, Appli
43	234.5	3.1	1187	3	US-08-545-860D-28	Sequence 28, Appli
44	234.5	3.1	1187	5	PCT-US94-04496-28	Sequence 28, Appli
45	234.5	3.1	1210	1	US-08-320-559-26	Sequence 26, Appli

#### ALIGNMENTS

RESULT 1  
US-08-891-640-2  
Sequence 2, Application US/08891640

Patent No. 6268173

GENERAL INFORMATION:

APPLICANT: Chamblon, Pierie

APPLICANT: Gronemeyer, Hinrich

APPLICANT: Voegel, Johannes

APPLICANT: Iutz, Yves

TITLE OF INVENTION: Transcriptional Intermediary Factor-2

NUMBER OF SEQUENCES: 14

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.

STREET: 1100 New York Avenue, NW, Suite 600

CITY: Washington

STATE: DC

COUNTRY: USA

ZIP: 20005-3934

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/891,640

FILING DATE: Herewith

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 60/021,247

FILING DATE: 12-JUL-1996

ATTORNEY/AGENT INFORMATION:

NAME: Steffe, Eric K.

REGISTRATION NUMBER: 36,688

REFERENCE/DOCKET NUMBER: 1383.0130001/EKS

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-371-2600

TELEFAX: 202-371-2540

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 1464 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-891-640-2

Query Match 100.0%; Score 7631; DB 4; Length 1464;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1464; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MSGMGENTSDPSRAETKRRKRECPDOLGSPKRNTERKRNQENKKTIEELAEITAFNFNDI	60
Db	1	MSGMGENTSDPSRAETKRRKRECPDOLGSPKRNTERKRNQENKKTIEELAEITAFNFNDI	60
Qy	61	DNFNFKPDCKALKEITVKQIRIQIKQEOKAAANIDEVOKSDVSTGQGVYIDKDALGPMML	120
Db	61	DNFNFKPDCKALKEITVKQIRIQIKQEOKAAANIDEVOKSDVSTGQGVYIDKDALGPMML	120
Qy	121	EALDGFFVYVNLBEGNVFVSENVTOYLRTYNOEELMKSVYSLIHGDHTEFPKMLPSSI	180
Db	121	EALDGFFVYVNLBEGNVFVSENVTOYLRTYNOEELMKSVYSLIHGDHTEFPKMLPSSI	180
Qy	181	VNGSWMSGEPPRRNSHTFNCRLMYKPLPDSEEGHNOHAHQYETMOCFAVSQPKSTKE	240
Db	181	VNGSWMSGEPPRRNSHTFNCRLMYKPLPDSEEGHNOHAHQYETMOCFAVSQPKSTKE	240
Qy	241	EGEDJLOSLICVARKYPMKERPLYPSSEFTTQDLOLQKITSLDTSTMKRAAMPQMEDJY	300
Db	241	EGEDJLOSLICVARKYPMKERPLYPSSEFTTQDLOLQKITSLDTSTMKRAAMPQMEDJY	300
Qy	301	RRCIOKTHAQHEGESVYAKRHHHEVYLROGLASQIYRPSLSDGTLVAAQTKSLIRSQT	360
Db	301	RRCIOKTHAQHEGESVYAKRHHHEVYLROGLASQIYRPSLSDGTLVAAQTKSLIRSQT	360
Qy	361	TNEPOLVYSLHMLHREONVCMNPDLTGOTGMCKPLPISNSPAHQALCSPMGQDMTIS	420
Db	361	TNEPOLVYSLHMLHREONVCMNPDLTGOTGMCKPLPISNSPAHQALCSPMGQDMTIS	420
Qy	421	SNINFPINGPEQMGMGKFFGGSGGNHNSGQATTPDGSNYALAKMNSPQSSPGMMPG	480
Db	421	SNINFPINGPEQMGMGKFFGGSGGNHNSGQATTPDGSNYALAKMNSPQSSPGMMPG	480
Qy	481	QPTSMLSPRHRMSPGVAGSPRIPPSQSPAGSLHSPVGYCSTGNSHTNSLNALOAL	540
Db	481	QPTSMLSPRHRMSPGVAGSPRIPPSQSPAGSLHSPVGYCSTGNSHTNSLNALOAL	540
Qy	541	SEGGCVSLGSSLASPDLMKMNILQNSPVNMPPLSLKMGSLDKDCFGILGEPSEBTTQOA	600
Db	541	SEGGCVSLGSSLASPDLMKMNILQNSPVNMPPLSLKMGSLDKDCFGILGEPSEBTTQOA	600
Qy	601	ESSCHPEQOKETNDPNLPVAVSSERADGQSLDKSGQTKLQLLTTKSKDQMEPPLASS	660
Db	601	ESSCHPEQOKETNDPNLPVAVSSERADGQSLDKSGQTKLQLLTTKSKDQMEPPLASS	660
Qy	661	LSDTNKSSTGSLPGSGSTHGTSLKEKHILHRLLDQSSSPVDLAKLTAEATGKDLQSSSS	720
Db	661	LSDTNKSSTGSLPGSGSTHGTSLKEKHILHRLLDQSSSPVDLAKLTAEATGKDLQSSSS	720
Qy	721	STAQSEVYTIKQEVSPKKEKNALRLTLDKDOTKOIGPEITPKLERLDSKTDPAJSNTK	780
Db	721	STAQSEVYTIKQEVSPKKEKNALRLTLDKDOTKOIGPEITPKLERLDSKTDPAJSNTK	780
Qy	781	LIANKTEKEEMSFEPGQOPSELDNLEIILDDQNSQOLPOLPDTPPGAPASVYKQATI	840
Db	781	LIANKTEKEEMSFEPGQOPSELDNLEIILDDQNSQOLPOLPDTPPGAPASVYKQATI	840
Qy	841	NDLMLQTLAENSPLYTVGAQKTAALRISQSTFNNRPNRPGQLRLPNOMLPDITLQSPTAG	900
Db	841	NDLMLQTLAENSPLYTVGAQKTAALRISQSTFNNRPNRPGQLRLPNOMLPDITLQSPTAG	900
Qy	901	PEPPIRNSSPVSYIPQGMGNOGMIGNOGLNCSSTGMIGNSASAPTPMPSEMAPOSSA	960
Db	901	PEPPIRNSSPVSYIPQGMGNOGMIGNOGLNCSSTGMIGNSASAPTPMPSEMAPOSSA	960
Qy	961	VRYVCAATTSAMNRPVOGGMIRNPPASIPRPPSSQOGQOTLOSQVMNIGPSELPMNNGG	1020
Db	961	VRYVCAATTSAMNRPVOGGMIRNPPASIPRPPSSQOGQOTLOSQVMNIGPSELPMNNGG	1020
Qy	1021	POYSQQAAPPNOTAPWPESTILPIDQASFASQNSQPPGSSPDLLDTCHPAAESPDEGALL	1080
Db	1021	POYSQQAAPPNOTAPWPESTILPIDQASFASQNSQPPGSSPDLLDTCHPAAESPDEGALL	1080
Qy	1081	DOLYIALRNEGDLEIDRALGIPELYQSQOAVDPDEPSSQSDSNIMLEQAKAYFPOQVASQ	1140

```

Db      1081  DQVLIALRNFDEGEIDRALGIDPELVSSQAVDPEQFSSQSDSNIMLEQKAPVPEQYASQ  11400
QY      1141  AQMAQGSYSPKQDPNFHTMGQRFPSYATLRKQPPRGLRPGLVQVONQDLRLDLQHRLOAQ  12000
Db      1141  AQMAQGSYSPKQDPNFHTMGQRFPSYATLRKQPPRGLRPGLVQVONQDLRLDLQHRLOAQ  12000
QY      1201  QNRQPLNMQJNSVNSVNLVTLRPQPPQAPINAQMLAQORREILNOHLRQRMHQOQOYQO  12600
Db      1201  QNRQPLNMQJNSVNSVNLVTLRPQPPQAPINAQMLAQORREILNOHLRQRMHQOQOYQO  12600
QY      1261  RTLMARQGLNMPFSVNAVPSGMPATSNPPIPOANAQQPFPFNYSIQQDPGFTGATT  13200
Db      1261  RTLMARQGLNMPFSVNAVPSGMPATSNPPIPOANAQQPFPFNYSIQQDPGFTGATT  13200
QY      1321  POSPLMSPRMAHTQSPPMQQSQANPAYQAPSDINGNAQGNMGNSKFSQSPPHFGQAN  13800
Db      1321  POSPLMSPRMAHTQSPPMQQSQANPAYQAPSDINGNAQGNMGNSKFSQSPPHFGQAN  13800
QY      1381  TSMYSNNKNTINVSMAATMGMSNNQQTGQISMTSVTSVTSGLSMBGEQVNDPALRG  14400
Db      1381  TSMYSNNKNTINVSMAATMGMSNNQQTGQISMTSVTSVTSGLSMBGEQVNDPALRG  14400
QY      1441  NLFPNQLPGMDMKQEGDITTRKCY  1464
Db      1441  NLFPNQLPGMDMKQEGDITTRKCY  1464

RESULT      2
US-08-891-640-3
: Sequence 3, Application US/08891640
: Patent No. 6268173
: GENERAL INFORMATION:
: APPLICANT: Chambon, Pierre
: APPLICANT: Gronemeyer, Hinrich
: APPLICANT: Voegel, Johannes
: APPLICANT: Lutz, Yves
: TITLE OF INVENTION: Transcriptional Intermediary Factor-2
: NUMBER OF SEQUENCES: 14
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
: STREET: 1100 New York Avenue, NW, Suite 600
: CITY: Washington
: STATE: DC
: COUNTRY: USA
: ZIP: 20005-3934
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/891,640
: FILING DATE: Herewith
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 60/021,247
: FILING DATE: 12-JUL-1996
: ATTORNEY/AGENT INFORMATION:
: NAME: Steffe, Eric K.
: REGISTRATION NUMBER: 36,688
: REFERENCE/DOCKET NUMBER: 1383.0130001/EKS
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 202-371-2600
: TELEFAX: 202-371-2540
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1036 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: not relevant
: MOLECULE TYPE: peptide
:
US-08-891-640-3

```



Query Match 13.6%; Score 1037.5; DB 4; Length 1036;  
 Best Local Similarity 28.7%; Pred. No. 2,3e-66;  
 Matches 346; Conservative 179; Mismatches 339; Indels 343; Gaps 52;

```

OY 395 LNPISNSPBAQALCSGN--PGQDWLTSNINFPINGKREQ-----MGMPMGKF 441
DB 10 VNP--SIPAGVAVARSSTLPPSNMNVSTRIN-----RQSSDLSHSSHSNSQSGF 61
OY 442 GSGGGMNVSGM-----QATTPOGSNYALKNNSPQSSPGMNPQPTSMLEPRHMSDQV 496
DB 62 GCSPROQLVAVNALNKGQASS--QSSKPSLNLNPPMEGIGISLAQ---FMSPRQVTSGL 117
OY 497 AGSPRIIPSOQSP--AGSLHSPYGVCS--TGNSHSYTSSNLALDALSEGHVSLGSLA 553
DB 118 ATRPMRPNNSFPNISTLSLSPVGMTSSACNNNNRYSNIPVTSLOGMNEGPNNVSGFSA 177
OY 554 SPDLKMGNIQNSPVNMPNPLSKMGLSDKCFGLYGEPEEGTQGAESSCHPGOKET- 612
DB 178 SPVLKQMSQNSPSRLNTP--AKAESKONKEIASLNLNEMI-----QSONSSDQKPLDSG 231
OY 613 ---NDPNLPPAVSSERADQSRHDSKQTKLLQLLTTKSD--QMBPSPLASSLD----- 663
DB 232 LLHNNDRLL-----SDGSKY--SQTSHKLVLQTLTTAEQQLRHADIDTCKDVLSC 281
OY 664 ---TKDSTGSLPGSGSTGTSTLKEKHKILHRLDSSSPVDLAKLTAEATGCKDLSQESS 721
DB 282 GTSNSASANSSSGSCPSHSSSLTARHKILHRLQ--GSPSDITLTSVPDKKD--SASTSV 339
OY 722 TAPG---SEVTKOEPYPPKKE---NALRYLDDKD-----TKDIGLPEITPLERLD 770
DB 340 SVTGOVQGNSSIKELDLASKKESKODHRLKLDKDEKDLSTYNLSLDYKVAVERKE 399
OY 771 SKTDPASTKLIAMKTEKEMSEFEGDQSGSELDNLEELDLQ--SOLPOLFPDTRPGA 829
DB 400 -QMPBCNTPPTPKATPEEIKLEAQSQFTADLDQDLPLTLEKAAQLPGLCETDRMDG 458
OY 830 PAGSVDKAIIINDLMQLTAENSPTVPGAOKTALRISQSTFNRRPGQLRLLPQNLPL 889
DB 459 AVTSV-----TIKS 467
OY 890 DITLQSPGAGPPPIRNSSPYSVLPORGMMGNQMGNGNLGNSGTMGNSASRPTM 949
DB 468 EITIKS-----ELLP-----ASLQSAIARP-- 487
OY 950 PSGEAPQSSAVRYTCAATTSAMNRPVCGMIRNPASIPMRSSQPGQRTQLQSOVNI 1009
DB 488 -----TSRLNR-----LP----- 495
OY 1010 GPSELEMMMGROYSQOAPRNOTAPRP--ESTLPIDQASFAQNRQPPGSSPDLLCFHP 1068
DB 496 ---ELELALIDNOFGQPGT--GDQIPMTNNVTAINQ---SKSEDCISSQDELCPPT 547
OY 1069 AAEPSDEGALLDOL--YIALRNFDELIEDRALGIPELVISOAVD--PROFSSQDSN- 1123
DB 548 TVEGRNDEKALLEGVLSTLSEKDELELADRALGIDIKLV--QGGGLDVLSSEFPQOATP 606
OY 1124 -TMEQKAPVEPQOYASQOQAQSGSYSPMDPNFTMTGQRPSTYATLRMQPRP----- 1174
DB 607 PLIMEERENLYSQPYSSPPTAN-----LPSPQGMVRQKPSLGMVQVVPTRGAFSPG 661
OY 1175 -GLAPRTGLVQ--QPNLRLOLQHLQAO-----QNRQPLANOISNVNLTLPFG- 1223
DB 662 MGMPROTINRPPAPRPNQRLQLOQOOLQOQOOLQHNQOALINQPAAPVGINRRSQ 721
OY 1224 --VTQAPINQMLAQRQREILNQLROROHQOQOQOQOQOQOQOQOQOQOQOQOQOQO 1279
DB 722 QOITRPPRLNOMLAQRRELXSOOHROROLIQ---QORALIMQOQSGNNLPPS----- 773
OY 1280 SGMPATMSNPRIPOANAOQPPRPYGISQOPDPEGTG---ATTPOSPL----- 1325
DB 774 SGLPVQOTGNPRLPOGAPOQFPYRPYGTNPGRPASTSPFSQLANPEASLANRNSVSR 833

```

```

OY 1326 -----MSPRBAHT-----OSPM 1337
DB 834 GMTGINGGQFGTGILNIPQMOQNVQYPGAGMPQGEANFAPSLPSSSNVMPDIPPOSL 893
OY 1338 MOOSQANPAYQAPSDINGNAQMGNSMFSQ--QSPHFGQOATMSYNSNNINVSMA 1395
DB 894 LQOTPPASQOSP--DMKMQGALIGNNVFSQAQVQNP---TPADPGY--NNMSTVSM 948
OY 1396 TMTGMSNNQMTGOISMTSVTSSTGLSNGPEQVNDPALRGNLPEPNOLPGMDMIQ 1455
DB 949 GQNTVQNMNPPMAQOM-----SLQMPGMVTCPPQINDPALRHGLVCNQLSTDLKT 1005
OY 1456 EGDTRK 1462
DB 1006 EADGTQ 1012

```

RESULT 3  
 US-08-885-291-55  
 ; Sequence 55, Application US/08885291A  
 ; Patent No. 6057125  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Takahashi, Joseph S.  
 ; APPLICANT: Turek, Fred W.  
 ; APPLICANT: Pinto, Lawrence H.  
 ; TITLE OF INVENTION: CLOCK GENE AND GENE PRODUCT  
 ; FILE REFERENCE: 0290-5  
 ; CURRENT APPLICATION NUMBER: US/08/885,291A  
 ; EARLIER FILING DATE: 1997-06-30  
 ; EARLIER APPLICATION NUMBER: 08/816,693  
 ; NUMBER OF SEQ ID NOS: 55  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 55  
 ; LENGTH: 846  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; US-08-885-291-55

Query Match 4.2%; Score 317; DB 3; Length 846;  
 Best Local Similarity 21.5%; Pred. No. 2e-14;  
 Matches 176; Conservative 129; Mismatches 316; Indels 196; Gaps 39;

```

OY 31 KNTTEKRNDEQNKYTELELIFANFNDIDNFKRDPKCAILKETVQOIROIKQEKAA 90
DB 39 RKKSEKRRDQFNVLKELGSMLEPNAR-----KMDKSVLQKSIDFLR--KKHEITA 89
OY 91 AANIDEVQKSVSSGQGVYDKDALGPMLELDGFFVNVNLEGVNVVSENVQYLKRN 150
DB 90 OSDASEI--RQDKKPT--FLSNEEFTQLMLELDGFFLAIMDGSIIYVSESVTSLLEHL 145
OY 151 QBELNKKSVSYSLIAGDHTFEFNKLLPKSIYVGSMSGCEPR--RSHNFFNCML----- 203
DB 146 PSDLVQDSIFNTPRGHSESVYK--ILSTHLESDBSLTEPYLKSQKQLECCIMLGTIDP 204
OY 204 -----VKPLPDSSEEGHDN-----QEAHQ--KYETMQCFA-----VSQPKSIK 239
DB 205 KEPSTYEVYKFIKGNKSLNSVSSAHNGFEQTIQRTHRPSYEDRVCFAVATLAPQFIK 264
OY 240 EEGEDLOSLCLVAVRVPKKEKRPVLPSSSEFTTRQDLOGKITSLDSTIMRAMKRGWEDL 299
DB 265 E-----MCTV-----EEP-----NEEFTSRSLSEWKFLELD--HRAPITIGYLPF 302
OY 300 VARCIOKFAHQHEGESVYAKRHHEVLRQGLAFQIYRFSLSDGTLVAAQTKSKLIRSQ 359
DB 303 EVLGTSYDYIYVDDLENAKCHEN--LMQYKKGKCYRFLTKGQOWIMLQNHYYITYHQ 361
OY 360 TTNEPOLVLSLMLHREONVCYMPNDLTGQITGKPLNPISNSPAHQALCSGNPGQDML 419
DB 362 WNSRPEEIVCTHTVVSVAEVAERREELG-----IEESLPETAADKSDQSSDNR 412

```

```

QY 420 SSINFPINGPKQMGMPKRGCGSGMNHVSGMQATTPOGSNYALKMNS-PSQSSPGMN 478
DB 413 NT-----VSLKALERFDHS-----PTPSASRRSRKSSHTAVSPSSST 451
QY 479 PCQ-PTSMLSPRHRMSPGVAGSPRIPPSQFSPAGSLHSPVGVCSSTGNSSHTNSLNL 537
DB 452 PTKIPPTDTS-----PPRQHLPA---HEKM-----VORSSFSQSINS- 487
QY 538 QALSEGHVSLGSSSLASPDLMKGNLONSPVNMNPPPLS-----KMGSLDS-KDCFGLYG 590
DB 488 -----QSVGSSLTQP--VMSQATNLPi-----PQMSQFQFSAQIGAMQHLKD----- 528
QY 591 EPSEGTGAESSCHPEQKETNDPMLPPAVSERADG-OSRLHDSK-----GQTKLLOL 644
DB 529 -OLEQRTRMIEANIH-RQDELR--KIQEOLQWVGQGLQMFLOQSNPGLNFGSVQSSG 584
QY 645 LTTKSQMPESPPLASSLSPTNKDSTGSLPGS--GSTHGTSLKEKHKLHLRLDQSSSPVDL 703
DB 585 NSSNIQOLAPINNQGVPVPTNQIOSGMNTGHIGTT-----QHMTQQOQTLOSTSTQSOQ 637
QY 704 AKLTAAATGKDLQSESSS--TAPGSEVTTIKQEPVSPKKENALLRYLLDKDRTKDIGLPE 761
DB 638 NVLSGHSQOTSLPSQOTSTIAPLYNTMVISQPAAGSM-----VQIPS 680
QY 762 ITPKLERLDSKTPASNTKLIAMKTEKESSEFPDQ 798
DB 681 SMPQ-----NSTQSAAVTTF---TODROIREFSQGOO 708

```

RESULT 4  
US-09-107-847-2  
Sequence 2, Application US/09107847

```

PATENT INFORMATION:
APPLICANT: DICKMORTH, DAVID
APPLICANT: MICHALOVICH, DAVID
TITLE OF INVENTION: NOVEL USE
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Ratner & Prestia
STREET: P.O. Box 980
CITY: Valley Forge
STATE: PA
COUNTRY: USA
ZIP: 19482
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/107, 847
FILING DATE: 30-JUN-1998
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP 97304996.8
FILING DATE: 08-JUL-1997
ATTORNEY/AGENT INFORMATION:
NAME: Prestia, Paul F
REGISTRATION NUMBER: 23,031
REFERENCE/DOCKET NUMBER: GH-30003
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-407-0700
TELEFAX: 610-407-0701
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 846 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-107-847-2

```

Query Match 4.2%; Score 317; DB 3; Length 846;  
Best Local Similarity 21.5%; Pred. No. 26-14;  
Matches 176; Conservative 129; Mismatches 316; Indels 196; Gaps 39;

```

QY 31 KRTEKRNQENKYTELELAELIFANFNIDNENFPDKALLKETVKQIQIKDEKKA 90
DB 39 RNSSEKRRQDFNVLLIKEIGSMPLGNAR-----KMDKSTVLDQKSIDFLR--KHKEIFA 89
QY 91 AANIDEVQKSDSVSTQGVIDKDALPMMLEALDGFEEFVNVNEGNAVVFSENVTOYLRYN 150
DB 90 QSDASEI-RQDWKPT--FLSNEEPTQLMLEALDGFELAMTQDSITVYSESTSLLEHL 145
QY 151 QEELMKSVYSILHVDHTEFVKNLLPKSTVNGSGSGEPPR-RNSHTNCRML----- 203
DB 146 PSDLVQOSTIFNFTPEGEHEVYK-ILSTHLLBEDSLTPRELYSKNDLDFCCHNLKRTIDP 204
QY 204 -----VKPLPDSEEGHDN-----QEAHQ-KYETMGCA---VSQPKSIK 239
DB 205 KEPSTVEYVKEIFGNFKLSNSVSSAHNNGEGTIOPTHRRPSYEDRVCFVATVRLATPQFIK 264
QY 240 EEBEDLDQSLICVARRVPKKEPVLPSSSEFTTRQDLOKIKTSLDSTVRANKPMWEDL 299
DB 265 E-----MCTV-----EEP---NEEFTSRHSLKMFLELD--HRAPPIIGYLP 302
QY 300 VRRCIQFHAQHEGESVYAKRHHNEVLROGLAFSOLYRFSLSDGTLLVAQTKSKLIRSQ 359
DB 303 EVLGTSQDYIYHVDLLENLAKCHEH-LMQYKKKSCGYRFLRFGQGMWMLQTHYITTHQ 361
QY 360 TTNEPOLVITSLHMLREQNVCVNPNPLTGOTMGKPLNPJSSNSPAHQALCSGNPGQDML 419
DB 362 WNSRPEFIVCTHTVVSYAERARREELG-----IEESLPETAAKSDQSGSDNRI 412
QY 420 SSINFPINGPKQMGMPKRGCGSGMNHVSGMQATTPOGSNYALKMNS-PSQSSPGMN 478
DB 413 NT-----VSLKALERFDHS-----PTPSASRRSRKSSHTAVSPSSST 451
QY 479 PCQ-PTSMLSPRHRMSPGVAGSPRIPPSQFSPAGSLHSPVGVCSSTGNSSHTNSLNL 537
DB 452 PTKIPPTDTS-----PPRQHLPA---HEKM-----VORSSFSQSINS- 487
QY 538 QALSEGHVSLGSSSLASPDLMKGNLONSPVNMNPPPLS-----KMGSLDS-KDCFGLYG 590
DB 488 -----QSVGSSLTQP--VMSQATNLPi-----PQMSQFQFSAQIGAMQHLKD----- 528
QY 591 EPSEGTGAESSCHPEQKETNDPMLPPAVSERADG-OSRLHDSK-----GQTKLLOL 644
DB 529 -OLEQRTRMIEANIH-RQDELR--KIQEOLQWVGQGLQMFLOQSNPGLNFGSVQSSG 584
QY 645 LTTKSQMPESPPLASSLSPTNKDSTGSLPGS--GSTHGTSLKEKHKLHLRLDQSSSPVDL 703
DB 585 NSSNIQOLAPINNQGVPVPTNQIOSGMNTGHIGTT-----QHMTQQOQTLOSTSTQSOQ 637
QY 704 AKLTAAATGKDLQSESSS--TAPGSEVTTIKQEPVSPKKENALLRYLLDKDRTKDIGLPE 761
DB 638 NVLSGHSQOTSLPSQOTSTIAPLYNTMVISQPAAGSM-----VQIPS 680
QY 762 ITPKLERLDSKTPASNTKLIAMKTEKESSEFPDQ 798
DB 681 SMPQ-----NSTQSAAVTTF---TODROIREFSQGOO 708

```

RESULT 5  
US-09-496-672-55

```

Sequence 55, Application US/09496672
Patent No. 6291429
GENERAL INFORMATION:
APPLICANT: Takahashi, Joseph S.
APPLICANT: Turek, Fred W.
APPLICANT: Pinto, Lawrence H.
TITLE OF INVENTION: CLOCK GENE AND GENE PRODUCT
FILE REFERENCE: 0290-5

```

CURRENT APPLICATION NUMBER: US/09/496,672  
CURRENT FILING DATE: 2000-02-03  
PRIOR APPLICATION NUMBER: 08/885,291  
PRIOR FILING DATE: 1997-06-30  
PRIOR APPLICATION NUMBER: 08/816,693  
PRIOR FILING DATE: 1997-03-13  
NUMBER OF SEQ ID NOS: 55  
SOFTWARE: Patent In Ver. 2.0  
SEQ ID NO 55  
LENGTH: 846  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-496-672-55

Query Match 4.2% Score 317; DB 4; Length 846;  
Best Local Similarity 21.5%; Pred. No. 2e-14;  
Matches 176; Conservative 129; Mismatches 316; Indels 196; Gaps 39;

QY 31 KRTEKRRREOENKYEIEELALIFANFNIDNPNFKPKCALIKETVQIQIRKEOEKAA 90  
DB 39 RNMSEKRRRODFNVILIKELGSMFGNAR-----KMDKSTVLOKSIDFLR--KHKEITTA 89  
QY 91 AANIDEVOKSDVSTGGGVIDKDALGPMMLBALDGFEEVNVLEGNVVFSENVTOYLRYN 150  
DB 90 QSDASEI-RODMKPT---FLSNEEFTQMLEALDGFIAIMTDSIIYVSEVSTLLEHL 145  
QY 151 QEELMNSVSVILVGDHTEFEVKMLPKRSYVNGSWSGEPPR-RNSHTFNCRL----- 203  
DB 146 PSDLVDSIFNFIPEGHESEYK--ILSTHLLSDSLTPEYLKSNQLEFCCHMLRGITDP 204  
QY 204 -----VKPLPDEEGHND-----QEAHQ-KYETMOGFA-----VSOPKSIK 239  
DB 205 KEPSTYEYKFIKFNKLSNVSSAHNGFEGTIOGTHRPSTEDVNCFAIVRLATPOGK 264  
QY 240 EEGEDLOSLCIVARRVPMKERPVLPSESEFTTQDLOGKITSLDTSTMARAMPGEWDL 299  
DB 265 E-----MCTV-----EEP-----NEEFTSRHSLEMKFLFLD---HRAPPIIGYLPF 302  
QY 300 VRRCIOGFHAGHESESVYAKRHHHEVLRQGLAFSQIYRESLSGTLVAOTKSKLIRSQ 359  
DB 303 EVLGTSGDYVYHVDLELACHEH-LMQYKGKSCYYRFLTKGOQWMTLQTHYYITYHQ 361  
QY 360 TTNEPOLVLSLHMLHREQNCVNMNPDLTGOTMGKPLNPISNSPAHQALCSGNPGDMLT 419  
DB 362 WNSRPERIVCHYVSYAEVAERRELG-----IESLSLETADKQSDGSDNRI 412  
QY 420 SSNINPDPKQMGMPMGRFGSGGMNHYSGMATTPOGSNYALKNNS-PSQSSPGMN 478  
DB 413 NT-----VSLKEALERFDHS-----PTPSASSRSRSKSSHTAVSDPST 451  
QY 479 PGQ-PTSMSPRRHMSGCVAGSPRIIPSOQSPAGSLSPVCGSTGNSHTYSSLNAL 537  
DB 452 PTKIPPTST-----PPROHLPA---HEKM-----VORRSFSSOSINS- 487  
QY 538 QALSEGHVGLSSGLASPDLMGNLONSPVNMNRPIS-----KMGSLDS-KDQFGLYG 590  
DB 488 -----QSVGSSLTP---VMSQATNLP---POGMSQFOFSQDLGAMOHKLD----- 528  
QY 591 EPSEGTGAESSCHPEOKETNDPNLPRVASSERADG-QSRLHDSK-----GOTKLLOL 644  
DB 529 -QLEQPRIMEANIH-RQOEELR--KIOEQLOMWHGGGLQMFLOOSNGULFGSVQLSSG 584  
QY 645 LTTKSDOMESPPLASSLSDTNKSDTSLPGS-GSTHGTSLKEKHKLHRLLODSSSPVDL 703  
DB 585 NSSNIQDLAIPINMOGVVPTWLOISGMNTHGIGTT-----QHMIOQOOSTOSTSOQ 637  
QY 704 AKLAEATGKDLQSESS--TAPGEVTLKOEVPYSPKKKENALLRYLLDKDXTDIDGPE 761  
DB 638 NVLSGHQOOSTSLPSQOSTLAPLNTWVLSQAPAGSM-----VOIIPS 680  
QY 762 ITPKLERIDSKTDPASNTKLIAKTEKEEMSFEPGDQ 798  
DB 762 ITPKLERIDSKTDPASNTKLIAKTEKEEMSFEPGDQ 798

DB 681 SMPQ-----NSTOSAAVTF-----TODROIRFSOGQO 708

## RESULT 6

US-08-816-693A-2  
Sequence 2, Application US/08816693A  
Patent No. 5874241  
GENERAL INFORMATION:  
APPLICANT: Takahashi, Joseph S  
APPLICANT: Turek, Fred W  
APPLICANT: Pinto, Lawrence H  
TITLE OF INVENTION: Clock Gene and Gene Product  
NUMBER OF SEQUENCES: 53  
CORRESPONDENCE ADDRESS:  
ADDRESS: Dressler, Rocky, Milanov & Katz  
STREET: Two Prudential Plaza, Suite 4700  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60601  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/816,693A  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: No. 5874241thrup, Thomas E  
REGISTRATION NUMBER: 33,268  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-616-5400  
TELEFAX: 312-616-5460  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 855 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-816-693A-2

Query Match 4.1% Score 309.5; DB 2; Length 855;  
Best Local Similarity 21.7%; Pred. No. 7e-14;  
Matches 150; Conservative 107; Mismatches 254; Indels 181; Gaps 27;

QY 31 KRTEKRRREOENKYEIEELALIFANFNIDNPNFKPKCALIKETVQIQIRKEOEKAA 90  
DB 39 RNMSEKRRRODFNVILIKELGSMFGNAR-----KMDKSTVLOKSIDFLR--KHKEITTA 89  
QY 91 AANIDEVOKSDVSTGGGVIDKDALGPMMLBALDGFEEVNVLEGNVVFSENVTOYLRYN 150  
DB 90 QSDASEI-RODMKPT---FLSNEEFTQMLEALDGFIAIMTDSIIYVSEVSTLLEHL 145  
QY 151 QEELMNSVSVILVGDHTEFEVKMLPKRSYVNGSWSGEPPR-RNSHTFNCRL----- 203  
DB 146 PSDLVDSIFNFIPEGHESEYK--ILSTHLLSDSLTPEYLKSNQLEFCCHMLRGITDP 204  
QY 204 -----VKPLPDEEGHND-----QEAHQ-KYETMOGFA-----VSOPKSIK 239  
DB 205 KEPSTYEYKFIKFNKLSNVSTHNGFEGTIOGTHRPSTEDVNCFAIVRLATPOGK 264  
QY 240 EEGEDLOSLCIVARRVPMKERPVLPSESEFTTQDLOGKITSLDTSTMARAMPGEWDL 299  
DB 265 E-----MCTV-----EEP-----NEEFTSRHSLEMKFLFLD---HRAPPIIGYLPF 302  
QY 300 VRRCIOGFHAGHESESVYAKRHHHEVLRQGLAFSQIYRESLSGTLVAOTKSKLIRSQ 359  
DB 303 EVLGTSGDYVYHVDLELACHEH-LMQYKGKSCYYRFLTKGOQWMTLQTHYYITYHQ 361  
QY 360 TTNEPOLVLSLHMLHREQNCVNMNPDLTGOTMGKPLNPISNSPAHQALCSGNPGDMLT 419

```
Db 362 WMSRPEFIVCTHTVSYAEVRAERRELQ-----IEESLPETAADKSDQSDGNRI 412
Oy 420 SSINFPINGPKPEOMGMPRGSGGMNHVSGMATTPQGSNYALKMNS-PSQSSPGNN 478
Db 413 NT-----VSLKEALERDHS-----PTPSASSRSRKSSTHAIVADPSSST 451
Oy 479 PCQ-PTSMULSP-----RHMSPGVAGSPRIIPSOFPAGSLHSPVGVCSST 523
Db 452 PTKIPTDTSTPPROHLPAHEKMTQRRSSFSQSINSQVGPSTLPQAMSOAANLPI--PQ 509
Oy 524 GNSHSTNSLNAOLAISE-----GHGVSLGSSL 552
Db 510 GMSQFQFSAQLGAMQHLKDQLBQRTMRIBANIHROOEELRKIOEOLQVNGGGLQMLQ 569
Oy 553 ASPDLKMGNLQNS-----PYMNP--PLSKMSLDSKDCFGLYGPESEG----- 595
Db 570 SNGLNFSGVOLSSGNSNIQQLTPVNMGOVYPANOVOS-----GHISTGQHMIO 619
Oy 596 -----TTGQAESCHPGEOKETNDPILPPA 620
Db 620 QOTLOSTTQSQSVMSGSHSOOTSLPSQTPS 651
```

```
RESULT 7
US-08-885-291-2
; Sequence 2, Application US/08885291A
; Patent No. 6057125
; GENERAL INFORMATION:
; APPLICANT: Takahashi, Joseph S.
; APPLICANT: Turek, Fred W.
; APPLICANT: Pinto, Lawrence H.
; TITLE OF INVENTION: CLOCK GENE AND GENE PRODUCT
; FILE REFERENCE: 0290-5
; CURRENT APPLICATION NUMBER: US/08/885,291A
; EARLIER APPLICATION NUMBER: 08/816,693
; EARLIER FILING DATE: 1997-03-13
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 2
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Mus musculus
US-08-885-291-2
```

```
Query Match 4.1%; Score 309.5; DB 3; Length 855;
Best Local Similarity 21.7%; Pred. No. 7e-14; Indels 181; Gaps 27;
Matches 150; Conservative 107; Mismatches 254;
```

```
Oy 31 KRNTKRNREQENKYIEELAEILFANFNIDNENFRPDKCAILKETVQIQTKEQEKAA 90
Db 39 RNKSEKRRQGFVLKELGSLMPLGNAR-----KMDKSTVLOKSIDFLR--KHKETTA 89
Oy 91 AANIDEVQKSDVSTGOGVIDKDALGPMLEALDGEFFVNLGNAVVENVSENVTOYLRYN 150
Db 90 QSDASEI-RQDMKPT--FLSNEEFQOLMEALDGEFLAIMTDSIIVSEVTSLEHL 145
Oy 151 QEELMKSVSYILHVDHTEFVKNLLPKSVNGSGMGEPPR-RNSHTFCRML----- 203
Db 146 PSDLVOSTINFTPEGHSVYK-ILSTHLESDSLTPRELYLSKNQLEFCCHMLRGITDP 204
Oy 204 -----VKPLPDEEGHDN-----QEAHQ-KYETMOGFA-----VSQPKSIK 239
Db 205 KEPSTYEYVAFIGNFKSLTSTHNGFEGTQRTHRPEYEDRVCFVAIVRLATPQFIK 264
Oy 240 EEGEDLQSLICVARRVPMKERVPLDSSSEFTTRDLOGKITSLDSTYRAMKPGMEDL 299
Db 265 E-----MCTV-----NEEFTSRHSLEMFLELD--HRAPIILGYLPF 302
Oy 300 VRCICQFHAHQGESVYAKRRHHHEVLRGGLAFSQTYYRSLSDGTILVAQOTSKILRSQ 359
```

```
Db 303 EYLGTSGYTYHVDLEENLAKCHEN-LMOYKGKSCYYFELFKGOQWIMLOTHYITYHQ 361
Oy 360 TTNEPOLVLSLHMLHREQVNCVMPNDLGTQMGKPLNPLISSNSPAHQALCSGNGODMTL 419
Db 362 WMSRPEFIVCTHTVSYAEVRAERRELQ-----IEESLPETAADKSDQSDGNRI 412
Oy 420 SSINFPINGPKPEOMGMPRGSGGMNHVSGMATTPQGSNYALKMNS-PSQSSPGNN 478
Db 413 NT-----VSLKEALERDHS-----PTPSASSRSRKSSTHAIVADPSSST 451
Oy 479 PCQ-PTSMULSP-----RHMSPGVAGSPRIIPSOFPAGSLHSPVGVCSST 523
Db 452 PTKIPTDTSTPPROHLPAHEKMTQRRSSFSQSINSQVGPSTLPQAMSOAANLPI--PQ 509
Oy 524 GNSHSTNSLNAOLAISE-----GHGVSLGSSL 552
Db 510 GMSQFQFSAQLGAMQHLKDQLBQRTMRIBANIHROOEELRKIOEOLQVNGGGLQMLQ 569
Oy 553 ASPDLKMGNLQNS-----PYMNP--PLSKMSLDSKDCFGLYGPESEG----- 595
Db 570 SNGLNFSGVOLSSGNSNIQQLTPVNMGOVYPANOVOS-----GHISTGQHMIO 619
Oy 596 -----TTGQAESCHPGEOKETNDPILPPA 620
Db 620 QOTLOSTTQSQSVMSGSHSOOTSLPSQTPS 651
```

```
RESULT 8
US-09-496-672-2
; Sequence 2, Application US/09496672
; Patent No. 6291429
; GENERAL INFORMATION:
; APPLICANT: Takahashi, Joseph S.
; APPLICANT: Turek, Fred W.
; APPLICANT: Pinto, Lawrence H.
; TITLE OF INVENTION: CLOCK GENE AND GENE PRODUCT
; FILE REFERENCE: 0290-5
; CURRENT APPLICATION NUMBER: US/09/496,672
; CURRENT FILING DATE: 2000-02-03
; PRIOR APPLICATION NUMBER: 08/885,291
; PRIOR FILING DATE: 1997-06-30
; PRIOR APPLICATION NUMBER: 08/816,693
; PRIOR FILING DATE: 1997-03-13
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 2
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-496-672-2
```

```
Query Match 4.1%; Score 309.5; DB 4; Length 855;
Best Local Similarity 21.7%; Pred. No. 7e-14; Indels 181; Gaps 27;
Matches 150; Conservative 107; Mismatches 254;
```

```
Oy 31 KRNTKRNREQENKYIEELAEILFANFNIDNENFRPDKCAILKETVQIQTKEQEKAA 90
Db 39 RNKSEKRRQGFVLKELGSLMPLGNAR-----KMDKSTVLOKSIDFLR--KHKETTA 89
Oy 91 AANIDEVQKSDVSTGOGVIDKDALGPMLEALDGEFFVNLGNAVVENVSENVTOYLRYN 150
Db 90 QSDASEI-RQDMKPT--FLSNEEFQOLMEALDGEFLAIMTDSIIVSEVTSLEHL 145
Oy 151 QEELMKSVSYILHVDHTEFVKNLLPKSVNGSGMGEPPR-RNSHTFCRML----- 203
Db 146 PSDLVOSTINFTPEGHSVYK-ILSTHLESDSLTPRELYLSKNQLEFCCHMLRGITDP 204
Oy 204 -----VKPLPDEEGHDN-----QEAHQ-KYETMOGFA-----VSQPKSIK 239
Db 205 KEPSTYEYVAFIGNFKSLTSTHNGFEGTQRTHRPEYEDRVCFVAIVRLATPQFIK 264
Oy 240 EEGEDLQSLICVARRVPMKERVPLDSSSEFTTRDLOGKITSLDSTYRAMKPGMEDL 299
```

```

Db 265 E-----MCTV-----EEP-----NEEFTSRHSLMKFLFD---HRAPIIGYLPF 302
QY 300 VBRICQKFAHBEESVYAKRHHEVLRQGLAFSQIYRFLSDDTLVAQTKSLINSQ 359
Db 303 EVLTSGDYHYHVDLELACHEH-LMVGKSKSCYRFLTKGQOMWLQTHYITVHQ 361
QY 360 TTNEPOLVLSLMLHREONCVMPNEDLTGCTMGKPLNLISSNPAHQALCSGNPGQDML 419
Db 362 WNSRPEFLVCHTYVSTAEVRAERRELQ-----IEESLPETADKSDSDSMDRI 412
QY 420 SSNINFLPGRKQMGWMPGRFGSGGNHVSQMOATTPQGSNVALKXNS-PSQSSPGMN 478
Db 413 NT-----VSLKEALERFDHS-----PTPSASRSRSKSSHTAVDPSSST 451
QY 479 PGQ-PTSLSP-----RHRMSPGVASPRIPPSQFPASLSHPVCVCSST 523
Db 452 PTKLPDTSTPPRQHLPAHERKTORSSFSQSINSQSVPSLQPMASQANLPI--PQ 509
QY 524 GNSHSYTSLSNALQALSE-----GHWVSLGSSL 552
Db 510 GMSQFQSAQAGAHQHLKQLEQRTMTAEANIHRQEEELRKIQEQLQVHKGGLQMLQ 569
QY 553 ASDPLKMGNIQNS-----PVNMP--PLSKMGLSKDCFGLYGEPSESG----- 595
Db 570 SNPLNGSVQLSSGNSNIQQLTPVNMGGQVVPANQVS-----GHISTGQHMIO 619
QY 596 -----TTGQAESCHAFGEOKETNDPMLPPA 620
Db 620 QOTLQSTSTQSOOSVMSGHSQOTSLPQSTPS 651

```

# RESULT 9

```

US-08-785-241-4
: Sequence 4, Application US/08785241
: Patent No. 5695963
: GENERAL INFORMATION:
: APPLICANT: McKnight, Steven L.
: APPLICANT: Tian, Hui
: TITLE OF INVENTION: Endothelial PAS Domain Protein
: NUMBER OF SEQUENCES: 7
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
: STREET: 268 BUSH STREET, SUITE 3200
: CITY: SAN FRANCISCO
: STATE: CALIFORNIA
: COUNTRY: USA
: ZIP: 94104
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/785,241
: FILING DATE: 17-JAN-1997
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: OSMAN, RICHARD A
: REGISTRATION NUMBER: 36,627
: REFERENCE/DOCKET NUMBER: UTSD:1229
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 343-4341
: TELEFAX: (415) 343-4342
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 870 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: peptide
: US-08-785-241-4

```

```

Query Match 3.8%; Score 292; DB 1; Length 870;
Best Local Similarity 20.6%; Pred. No. 1.3e-12;
Matches 217; Conservative 138; Mismatches 406; Indels 294; Gaps 46;

QY 10 DPSRAETRRKREKCPDOLGPPSPKRNTEKRNREDENYIEELA-ELIFANFDINDNFNRPD 68
Db 6 EKKRSSERRKE-----KSRDAARCRSRSETEVEFLAEHLPLPH-----SVSSHLD 52
QY 69 KCAILKETVQKIQIODEKKAANAANIDEYKSDVSTGGGVIDKALCPMLLEALDGEFF 128
Db 53 KASIMRLEISFLRTHKLLSYVCSENESEAEAD-----QOMNLUTKALEGFLA 100
QY 129 VVNLGNVVFSENYTOYLRYNOEELMKSVYSLIHDGHTFVFNKILPKSYVNGSGMSG 188
Db 101 VVTQGDMLFLESENISKPMGLTQVELTGHSITDFTHPCDHEIRENL--SKNNGSGR- 156
QY 189 EPPRRNSHTFNCRLMLVKPLDSEEGHNDQEAHQYETMQCPAVSQPSIKEEG----- 242
Db 157 -----KSKDMSTREDFFMKMKCTVTRNGRTVNLKSAITWYLV 193
QY 243 -----EDLQSLICVARRVPMKERVPLP-SSESFTTRDQLOK 279
Db 194 HCTGQVKYVNCPRPHNSLQGYKEPLSLCLIMCEPIQHPSHMDIPDLSKTYFLSRSHMDK 253
QY 280 ITSLSPTSTKRAAMKPGW--EDLVRCIQKFAHQHGESVYAKRHHEVLRQGLAFSQIY 337
Db 254 FYCYCD--RITELGYHPEELIGRAVEFYALDSENM--KSHQNLCTGQGVVSGQY 307
QY 338 RPSLSDGTLVAAQKTSKLIRSQTNQTNPOLVLSLH--MLHREONCVMPNEDLTGQTMGRP- 394
Db 308 RMLAKHGQVWLETOGYIYNPRNLQPOCINQVNVLSIEIENDVYFSDQI--ESLFRPH 366
QY 395 --LNPISNSPAHQALCSGNPGQDMLSSNINFLPGRKE-----QMGWPG----- 439
Db 367 LMANNSIFDSS-----GKG-AVSEKSNFLFTKLKEPEELAQPLAPPGDAITSL 414
QY 440 RFGSGGGMNHVSQMOATTPQGSNVALKXNSPS-QSSPGMNPCOPTSMLSPRHRMSPGAVG 498
Db 415 DFGNMFEESSAYGKAIIIPPSQPMATELRSHTQSEASLPAFTV-----PQAAA 464
QY 499 SPRIPSOFPSPAGSLHSPVGVCSSTGNSHVSNSLNLQALSEGHVSLGSSLASPLK 558
Db 465 POSTTPSATSSSS-----CSTPNSPEDYITSLDNLK-----LEVIKLAMDTE 510
QY 559 MGNLQNSPVNMPNPLSKMGLSKDCFGLYGEPSEGTGQAESCHAFGEOKETNDPMLP 618
Db 511 AKDCSTQTDFFNELDLETLAPY-----IPMDGEDFQLSPIC-PEERLAEINPQST 559
QY 619 PAVSSERADGOSRLHDSKQGT-----KLLQLITTSQDNEPPLASS 660
Db 560 PQ-----HCFSAMTNIPOPLAPVAPRHSFLLDKFQOOLESKTEPEENRPMSSI 607
QY 661 LSDTKDSTGSLP--GSGSTHGTSLKEKHKILHLLDQSSSPVQALKLTATGKDLSQ 717
Db 608 FFDAG--SKASLPRCCGQASTPLSSMGGRSN-----TQMPRPPLHFGPTKAAVG-DQRT 659
QY 718 ESSSTAPGSEVYIKQEPVSPKKRENALLRYLLDKDQTDIGL--PEI-TPKLERLDSKTD 774
Db 660 EFLGAAP-----LGPVSPPHVSTFKTR-----SAGFGARGPDVLSPAVVAL----- 702
QY 775 PASNTKLAMKTEKEMSEFEP--GDQGSSELDNIE-ELIDLDQNSQLPQLFPDTRPGAP 830
Db 703 --SNKILKLRQLEYEQAFOQLDSGDPGQSTSHLMKMKMLRGSGSCP-LMPD----- 753
QY 831 AGSVQKQAIINDLMQTLAENSPTVPVGAOKTALRTISQSTFNPNRGQGLRLLPNQNLPLD 890
Db 754 -----KPLSANVPNDKQTQ--NPMRG-LGH--PLRLHLPLP 783
QY 891 ITLQSPGTAGP-----FPPIRNSSPYSVIPQPMGNGQMGIGNOGLGNSGTMGINS 943
Db 784 --QPPSAISPGENSKSRFPQCYATQYQ-----DYSISSAKHVSGM 822

```

OY 944 ASRPTMPSGEMAPSSAVRVCATTSANRRPV0G 978  
 DB 823 ASRLGPFSESYLLPBLTRYDC-----ENVNPLYG 852

# RESULT 10

US-08-785-310A-8

; Sequence 8, Application US/08785310A

; Patent No. 5840532

; GENERAL INFORMATION:

; APPLICANT: McKnight, Steven L.

; APPLICANT: Russell, David W.

; TITLE OF INVENTION: Neuronal PAS Domain Protein

; NUMBER OF SEQUENCES: 8

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP

; STREET: 268 BUSH STREET, SUITE 3200

; CITY: SAN FRANCISCO

; STATE: CALIFORNIA

; COUNTRY: USA

; ZIP: 94104

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/785,310A

; FILING DATE: 21-JAN-1997

; CLASSIFICATION: 536

; ATTORNEY/AGENT INFORMATION:

; NAME: OSMAN, RICHARD A

; REGISTRATION NUMBER: 36,627

; REFERENCE/DOCKET NUMBER: UTSD:1226

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (415) 343-4341

; TELEFAX: (415) 343-4342

; INFORMATION FOR SEQ ID NO: 8:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 816 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: peptide

; US-08-785-310A-8

; Query Match 3.8%; Score 289; DB 2; Length 816;

; Best Local Similarity 18.3%; Pred. No. 2e-12;

; Matches 211; Conservative 151; Mismatches 360; Indels 432; Gaps 42;

OY 4 MGNSTDPSSAETRRKRECPDQGPSPKRTTEKRNREOEKRYTEELAEFLFANFNIDNF 63

DB 1 MDDEKDRARARSRNK-----SEKKRDQFVLLIKELSSMLPGTR----- 41

OY 64 NFKPDKCALIKETVQKIOIKEDOEKAAANIDEVOKSDVSTQGVJDKDALGPMLEAL 123

DB 42 --KMDKTYVLEKVIQLQ--KHNEVSAQTEICIDQDWKRS-----PLSNELFQIMLEAL 93

OY 124 DGEFFVNVNLEGNVENVSENVTOYLRNQEELMANKSVYSILHVDHTEFVKLLPKSTIVG 183

DB 94 DGEVIYVTTDGSITIVSDSITPLGLPADVMDQNLNLFPEQEHSESVYKILSSHMLVTD 153

OY 184 GSWSGEPFRNSHTFCRML---VKP-----LPDSEEGHON--- 217

DB 154 SPSPEFLKSDNLEFYCHLLRGLNRPETTYEYIKFVGNFRSYNNVPSPCNGCFDUTLS 213

OY 218 OEAHQYETWQCF-----AVSQPSIKEEGBDLQSLICVAVRPMKERVPLPSESPTTR 273

DB 214 RPHVPLVGLADVCIATVATPLATPOFLKE-----MCVA-----DEPL-----EETTSR 254

OY 274 ODLOGITSLDTSTMBAMKPGW---EDLVRCIOKFNHQBEGESVSYAKRHHNEVLROG 330

DB 255 HSELEWFLFD---HRAPPILGYLPPEVLGTSGYNYH-----DDELLARQHHLMOG 307

OY 331 LAFSQIYRFSLDGTLVAQOTSKILRSQTNBPOLVSLH-----MLREQNVCM 382

DB 308 KKKSCCYRFLTLGGQWIMWLQTHYIYTHOMNSKPERIVCTHVSADVVRERQELALE 367

OY 383 NPDLTGQTKPLNPLISSNSPAHQALCSGNPGQDMTLSSINFPINGPREQMGMPYRFG 442

DB 368 DP-----PTEAM----- 374

OY 443 GSGGMNHVSGMQATTPOGSNTYALKMNSPSQSSPGMNGOPTMNLSPRHMSGVAGSPRI 502

DB 375 -----HPSAVK-----EKDSLEPPQPFNAL-----DMGASGLPSS 405

OY 503 PPSQFSPAGSLSPVGCSSSTGNSHSTNSLNAQALSEGHGVSLGSSIASPLKMGML 562

DB 406 P-----SPASSSS-----SHKSHITAMS----- 424

OY 563 QNSPVNMNPPPLSKMGLSDSKDGFGLYGP-SEGTTGAQESSCHPEQKETNDPMLPAV 621

DB 425 -----BPTSTPKLMAENSTYALPRLATLPQELPV0G 456

OY 622 SSEPADGOSRLDSSK-----QTKLQLLTTKSDQMPPLAS-----SLSDINKOSTGSL 672

DB 457 LSOAATMPALHSSASCIDLTKQLLSLPQTGLQSPAPVTPQSAQSFQTKD----- 511

OY 673 PGSGSTHGTSLKHKILHRLLDSSSPVDLAKLFAEATGKDLQSSSTAGSEVYTIQ 732

DB 512 -----QLEQRTRILQ-----ANIRMOQ 528

OY 733 EPVSPKKENALLRYLLDKDKTDKIDGLPETPKLERLSDKTDPASTKLIAMKTEKEENS 792

DB 529 EELHAKIOEOLCLVQ-----DSNVQWF-LQQPAVSL 558

OY 793 FEFGDDPGSELNLEILDDLQNSQLPQLPDRCPAGASVDKQALLINDMLQTLAENP 852

DB 559 FSSTQRPAAQ-----QQLQO-----RPAAPS---QPLVNV----- 586

OY 853 VFPVGAQKTLARISQS-----TFNNRPQOLGRLLPNQNLPL-DITLQSPGAGPF 902

DB 587 -PPLQGITSTQVYTNQHLRESNVISAQGPKRPSQQLPASPASRLSSLSLPSQFSSSTASVL 645

OY 903 PPIRNSPYSVLPQPGMMQNGMIGNQLGNS-STGMIGNSASRPTMPSGEMAPSSAV 961

DB 646 PGLSLTITIAFPQO-----DSSQCPSPDFGHDRLQLRLLSQIPMMGSCDARPSSEV 700

OY 962 RVT-----CATTSAMNRPVGGMIRNPASIPMRPSSQPGQRTLQS 1004

DB 701 SRTGRQVKKYAAQGVMPFSPDHPNTSSASTPV---ILMGOAVLHPSFPASRPSPLD----- 753

OY 1005 QVANIGPSELEMMNGPQXSOQAPNPQTAP-----WPESTLPIQASFASSQNRQPFSS 1058

DB 754 -----PQAQOQPPRYLQAPTLHSEQSDSL-----LSITFSQOPOTLQY 793

OY 1059 SPDDLCLPHPAES 1072

DB 794 AATQSTPQPPRPS 807

RESULT 11

US-08-816-693A-53

; Sequence 53, Application US/08816693A

; Patent No. 5874241

; GENERAL INFORMATION:

; APPLICANT: Takahashi, Joseph S

; APPLICANT: Turek, Fred W

; APPLICANT: Pluto, Lawrence H

; TITLE OF INVENTION: Clock Gene and Gene Product

; NUMBER OF SEQUENCES: 53

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Dressler, Rocky, Milnamow & Katz

; STREET: Two Prudential Plaza, Suite 4700

```

: CITY: Chicago
: STATE: Illinois
: COUNTRY: USA
: ZIP: 60601
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/816,693A
: FILING DATE:
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: No. 5874241thrup, Thomas E
: REGISTRATION NUMBER: 33,268
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 312-616-5400
: TELEFAX: 312-616-5460
: INFORMATION FOR SEQ ID NO: 53:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 816 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: peptide
: US-08-816-693A-53

```

```

Query Match      3.8%; Score 288; DB 2; Length 816;
Best Local Similarity 18.6%; Pred. No. 2.3e-12;
Matches 214; Conservative 148; Mismatches 361; Indels 430; Gaps 43;

```

```

OY 4 MGEVTSRRAETRRKRCPCPOLGSPKRNTEKRNREDEKNTIEELAEIIFANFNIDINF 63
DB 1 MDEDEKDRAKRASRNK-----SEKKRRDQFNVLIKELSSMLPGNIR----- 41
OY 64 NFKPDKCALIKETVQIQRKEQKAAANIDEVOKSDVSSGTGQVIDKALGPMLEAL 123
DB 42 --KMDKITVLEKVGIFLQ--KHNEVSAQTEICDIDQDMKPS---FLSNEETQMLRAL 93
OY 124 DGEFFVNLEGNNVVFSENVTOYLRYNOEELMNSVSYILHVGDTHEVKNLLPKSIYNG 183
DB 94 DGFIVVTTDSITIVSDSTPRLGHLRADVMDQNLNLFPEQHESEYKIKLSHMLVTD 153
OY 184 GSNAGEPPRRKSHFNENML--VKP-----LPDSEBECHDN--- 217
DB 154 SPSEPELKSMDLDEFYCHILKGLSNPKKEFPYEVYIKFVGNFRSYNNVPSPCNGCFDNTLS 213
OY 218 QEAHQKYEYMQCF---AVSOPKSIKEGEDLQSLICVARRVPMKEKRPVLPSSESFTTR 273
DB 214 RPHAVPLKDVCFATAYVALTPQFLKE-----KCYA-----DEPL---EEFTSR 254
OY 274 QDLQKITSIDTSTWRAAMKPGW--EDLVRCIOKFHAOHEGESVYAKRHHEVLRQ 330
DB 255 HSLMKKFLFLD---HRAPPIIGYLPFEVLVGTSGVNYHI---DDLELLARHQHQLMGFG 307
OY 331 LAEQIYRFLSDGTVAQAQKSKILRSQTNREPOLVLSLH-----MLHREONCVW 382
DB 308 KKKSCCYRFLKGGQOWIWLQHYHYITYHQMNSKPEFIVCTSHSVSYADAVRERROELALE 367
OY 383 NPDLTGOTMGRPLNPISNSPAHQALCSGNPGQDWTLSNINFPINGKMGCMGMRFG 442
DB 368 DP-----PTEAM----- 374
OY 443 GSGGMNVHSGQATTPQGSNYALMKNSPQSGPGMNPQOPTSMSPRHMSPGVAGSPRI 502
DB 375 -----HPSAVK-----EKDSSLPEQPFPNAL-----DMGASGLPSS 405
OY 503 PPSQFSPAGSLHSPVGVCSSTGNSHSTYNNSSLMALQALSEHGVSGLGSLASPLDKMNL 562
DB 406 P-----SPASSR-----SSHKSSHT-----AMSE----- 425

```

```

OY 563 QNSPVNMNPPLSKMGSLDSKDCFGLYGPEPSEGTGQAESSCHPGROKETNDPNLPAYS 622
DB 426 -----PISTPKL-----MAENSTTALPRATTLPOELPVQGL 457
OY 623 SERADGQSRHDSKG---QTRILQLTTRKSDOMESPPLAS-----SLDSTNKDSTGSLP 673
DB 458 SQAATMPALHSSASODTLKQLLQSLPQTGASPPAPVQTGSAQFSMQTKD----- 511
OY 674 GSGSTHGTSLKAKKHILHRLDODSSPYDLAKTLAEATGKDSLQSESSSTAPGESEYTIQ 733
DB 512 -----OLEQRTIRILQ-----ANTRMOE 529
OY 734 PVSPPKKEKALLRYLLDKDQTKDIGLPETTPKLERLDSKTPDASNTKLIAMTEKEENSF 793
DB 530 ELHKTIQELQCLVQ-----DSNVQM-FIQDPVVSLSF 559
OY 794 EPGDQSELDLLEETLDDIQLNSQLPOLFPDTPGAPAGSYDKQAIINDLMQTLAENSFV 853
DB 560 SSIQRPAAQ-----QQLQQ-----RPAAPS---QPOLVYN----- 586
OY 854 TPVGAQKTAALRISQS-----TFNNPREGQLRLLPNOMLPL-DITLQSPTAGAPPP 903
DB 587 TPLAQOISTQVNTQNLHRESNVISAQGFKPMRSSQLLPASGRSLSLPSQFSSTASYLP 646
OY 904 PIRNSSPVSVIPQPMGMGQMGIGNQGNGNS-STGMIGNSASRPMTSGEAPQSSAVR 962
DB 647 PELSLTTTAPTPQ-----DSSQCQPSPDGHRQLRLLSQPIQIPMPEPSCDAPPSFVS 701
OY 963 VT-----CAATTSAMNRPVQGMIRNPASIPMRPSSQPGQROTLOSQ 1005
DB 702 RRGROVKYAAQSGVMPSPDSHTNSSASRPV---LLMGAVLHPEFPAASRPPLQ----- 753
OY 1006 VNNITPSELEMMNGPQYSSQQAAPPNOTAP-----WPESILPTQASFASONRQPFSS 1059
DB 754 -----PAQAQDQPPYLOAPTLHSEQDPDSL---LTFSSQPGTLGYA 794
OY 1060 PDDLCPHPAAS 1072
DB 795 ATQSTPPQPPPS 807

```

```

RESULT 12
US-08-885-291-53
: Sequence 53, Application US/08885291A
: Patent No. 6057125
: GENERAL INFORMATION:
: APPLICANT: Takahashi, Joseph S.
: APPLICANT: Turek, Fred W.
: APPLICANT: Pinto, Lawrence H.
: TITLE OF INVENTION: CLOCK GENE AND GENE PRODUCT
: FILE REFERENCE: 0290-5
: CURRENT APPLICATION NUMBER: US/08/885,291A
: EARLIER FILING DATE: 1997-06-30
: EARLIER APPLICATION NUMBER: 08/816,693
: NUMBER OF SEQ ID NOS: 55
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 53
: LENGTH: 816
: TYPE: PRT
: ORGANISM: Mus musculus
US-08-885-291-53

```

```

Query Match      3.8%; Score 288; DB 3; Length 816;
Best Local Similarity 18.6%; Pred. No. 2.3e-12;
Matches 214; Conservative 148; Mismatches 361; Indels 430; Gaps 43;

```

```

OY 4 MGEVTSRRAETRRKRCPCPOLGSPKRNTEKRNREDEKNTIEELAEIIFANFNIDINF 63
DB 1 MDEDEKDRAKRASRNK-----SEKKRRDQFNVLIKELSSMLPGNIR----- 41
OY 64 NFKPDKCALIKETVQIQRKEQKAAANIDEVOKSDVSSGTGQVIDKALGPMLEAL 123

```

```

Db      42  --KMDKITVLEKVIQFLQ--KHNEVSAQTEICIDQDMKPS-----FLSNEEFTQLMLEAL 93
OY      124 DGEFFVNNLEGNVVFSENVTOYLRYNOEELMKSVSYLIHVGDNHFEVNNLPRKSYNG 183
Db      94 DGFVIYVTTDGSITIVSDSTTPRLGHLRADVMQONLNLFLPBOHSHVYKILISSHMLYTD 153
OY      184 GWSGEPFRNRSHTFNCRLM--VKP-----LPDSEEGHND-- 217
Db      154 SPSPREFKSDNDELFYCHLRLGSLNPREPPTVEYIKFVGNFRSVNNVPSGNGFDTLS 213
OY      218 QEAHQKTEMQCF-----AVSQPKSIKEEGEDLOSCLICVARRYRMKERRVLPSESESTTR 273
Db      214 RCPHVPLGKDVCFIATVRLATPQFLKE-----MCVA-----DEPL-----EESTSR 254
OY      274 ODLOGKITSLDSTMRAMKPGW---EDLVRRCIOKFHAQHEGESVYAKRHHNEVLROG 330
Db      255 HSLFMKFLFD---HRAPIIGIYLPREVLTSGYNYHT---DDELLARCQIOMLMQFG 307
OY      331 LAFSQTIRFSLSDGTIVAQAOTKSKLIRSQTTNEPOLVISH-----MLHREQNYCVM 382
Db      308 KGKSCCYRFLTKGQOMIWLQTHYITGHQWNSKREFIVCHSVSVYADVAVERRQELALE 367
OY      383 NPDLTGOTMGKPLNPISNSNPAHQALCSGNPGQDMTSSINPFINPKQMGKMPGRFG 442
Db      368 DP-----PTKAM----- 374
OY      443 GSGGMNVHSGMOATTPQGSNYVALKMNSPQSPGMNPGOPTSMLSPRHMSPGVAGSPRI 502
Db      375 -----HPSAVK-----EKDSSLEPPOFPNAL-----DMGASGLPSS 405
OY      503 PPSQFSPAGSLHSPVGCSTGNSHSTYNSLNAQALSGHGVSLGSSLASPDLMKGNL 562
Db      406 P-----SPSASSR-----SSHKSST-----AMSE----- 425
OY      563 QNSPVNMNPPPLSKMGSLSKDCGFLYGEPSSEGTGOAESCHPGEOKETNDNPLPAVS 622
Db      426 -----PISTPTKRL-----MAENSTALPPRALPOLPQGL 457
OY      623 SERADQSRLDHDSKG---OTKLLOLLTTKSDOMEPSPLAS-----SLSDNTDSTGSLP 673
Db      458 SQAAITPTALHSSASCDLTQOLLQSLPQTGLOSPPAPVQFQSGFSPFTIKD----- 511
OY      674 GSGSTHGTSLKEKHILHRLLODSSSPVDLAKLTAETGKDLQOESSSTAAGESEVITKOE 733
Db      512 -----QEQRRRILO-----ANIRMQOE 529
OY      734 PVSPPKKENALLRYLDKDKDTKDIGLPETPKLERLDSKTDPRASNTKILMKTEKEMSF 793
Db      530 ELKHIOEQQLCIVQ-----DSNVQMF--LQOPPAVSLSF 559
OY      794 EPGDQPESELNLEILDDQNSOLPOLPDTGPRGAPAGSVQKQALINDLMQTLAENSVP 853
Db      560 SSTQIRAAQ-----QQLQO-----KRAAPS-----QQLVYVN----- 586
OY      854 TPVGAQKATLALISOS-----TENNPRPGOLGRLLPNONLPL--DITLQSPGAGPFP 903
Db      587 TPLQOQITSTQVINOHLRESNVIISAQGPKPMSSQLLPASGRSLSSLPQFSSTASVLP 646
OY      904 PINSSSYVITPQGMNGMGNMGICNGNLGNS--STGMIGNSASAPTPPSEMAPQSAVR 962
Db      647 PGLSLTTIATPQ-----DSDQCPSPDFGHQDLRLLSOPILQPMMPGSCDAAPSEVS 701
OY      963 VT-----CAATTSAMNRPVQGMIRNPASISIPMRSSPGQGHQTLQSQ 1005
Db      702 RTGRQVYKASQVWFSPBDHPITNSASTPV---LLMGQAVLHNSFPASRSPLO----- 753
OY      1006 VNIIGPSELEMMNGGPOYSQQAAPPNQTAP-----WPESILPTDAQSFASQNRQPGSS 1059
Db      754 -----PAOAOQOQPPPYLQAPTSLHSEQPDSIL-----LSTFSQOQGTGLGYA 794
OY      1060 PDOLLCPHRAES 1072

```

```

Db      795 ATOSTPPOPPRPS 807

RESULT 13
US-09-496-672-53
; Sequence 53, Application US/09496672
; Patent No. 6291429
; GENERAL INFORMATION:
; APPLICANT: Takahashi, Joseph S.
; APPLICANT: Turek, Fred W.
; APPLICANT: Pinto, Lawrence H.
; TITLE OF INVENTION: CLOCK GENE AND GENE PRODUCT
; FILE REFERENCE: 0290-5
; CURRENT APPLICATION NUMBER: US/09/496,672
; CURRENT FILING DATE: 2000-02-03
; PRIOR APPLICATION NUMBER: 08/885,291
; PRIOR FILING DATE: 1997-06-30
; PRIOR APPLICATION NUMBER: 08/816,693
; PRIOR FILING DATE: 1997-03-13
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 53
; LENGTH: 816
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-496-672-53

Query Match      3.8% Score 288; DB 4; Length 816;
Best Local Similarity 18.6%; Pred. No.2.3e-12;
Matches 214; Conservative 148; Mismatches 361; Indels 430; Gaps 43;

OY      4  MGENTSDPSRAETRRKRECPDQLGSPKRTKERNRQENKYLEELAEILFANFNIDNF 63
Db      1  MDEDEKRAKASRNK-----SEKKRRQGFVLLIKELSSMLPGNIR----- 41
OY      64 NEKPKDCALIKETVQKIQRIKEOEKAAANIDEVQXSDVSTQGVYDKDALCPMMEAL 123
Db      42  --KMDKITVLEKVIQFLQ--KHNEVSAQTEICIDQDMKPS-----FLSNEEFTQLMLEAL 93
OY      124 DGEFFVNNLEGNVVFSENVTOYLRYNOEELMKSVSYLIHVGDNHFEVNNLPRKSYNG 183
Db      94 DGFVIYVTTDGSITIVSDSTTPRLGHLRADVMQONLNLFLPBOHSHVYKILISSHMLYTD 153
OY      184 GWSGEPFRNRSHTFNCRLM--VKP-----LPDSEEGHND-- 217
Db      154 SPSPREFKSDNDELFYCHLRLGSLNPREPPTVEYIKFVGNFRSVNNVPSGNGFDTLS 213
OY      218 QEAHQKTEMQCF-----AVSQPKSIKEEGEDLOSCLICVARRYRMKERRVLPSESESTTR 273
Db      214 RCPHVPLGKDVCFIATVRLATPQFLKE-----MCVA-----DEPL-----EESTSR 254
OY      274 ODLOGKITSLDSTMRAMKPGW---EDLVRRCIOKFHAQHEGESVYAKRHHNEVLROG 330
Db      255 HSLFMKFLFD---HRAPIIGIYLPREVLTSGYNYHT---DDELLARCQIOMLMQFG 307
OY      331 LAFSQTIRFSLSDGTIVAQAOTKSKLIRSQTTNEPOLVISH-----MLHREQNYCVM 382
Db      308 KGKSCCYRFLTKGQOMIWLQTHYITGHQWNSKREFIVCHSVSVYADVAVERRQELALE 367
OY      383 NPDLTGOTMGKPLNPISNSNPAHQALCSGNPGQDMTSSINPFINPKQMGKMPGRFG 442
Db      368 DP-----PTKAM----- 374
OY      443 GSGGMNVHSGMOATTPQGSNYVALKMNSPQSPGMNPGOPTSMLSPRHMSPGVAGSPRI 502
Db      375 -----HPSAVK-----EKDSSLEPPOFPNAL-----DMGASGLPSS 405
OY      503 PPSQFSPAGSLHSPVGCSTGNSHSTYNSLNAQALSGHGVSLGSSLASPDLMKGNL 562
Db      406 P-----SPSASSR-----SSHKSST-----AMSE----- 425
OY      563 QNSPVNMNPPPLSKMGSLSKDCGFLYGEPSSEGTGOAESCHPGEOKETNDNPLPAVS 622

```



```

Db 426 -----DSTPTKL-----MAENSTALPRPATLPOELPVGCL 457
QY 623 SERADGOSRLDMSG-----OTKLOLITTKSDQMPSPSLAS-----SLSDNKNKSTGSLP 673
Db 458 SOAATMPALSSASCDLTTLKOLLSLPOTGLSPAPVTFPSAOFMSFOITKD-----511
QY 674 GSGSTHGTSLKEKHILHRLLDSSSPVDLAKLTAATGKGLDSSESTAGSEVITKOE 733
Db 512 -----OLEQRTIRILO-----ANIRMOOE 529
QY 734 PVSEKKENALRYLKDQDTRKIDGLPEITPKLERLDSKTPASNTKLIAMKTEKEEMSF 793
Db 530 ELKHIOEOLCLVQ-----DSNVQMF-LQQPAVLSLF 559
QY 794 EPQGPSELDNLEIIDDLDLNSQLPOLFPPTRGAPAGSDKQAIINDLMQLAENSVP 853
Db 560 SSIGRPAAQ-----QDLQO-----RPAAPS--QPOLVNV-----586
QY 854 TPVGAOKTALRISOS-----TFNPRPGOLGRLLPNOLPL-DITLQSPGAGPFP 903
Db 587 TPLGQITSTQVTHQHLRESNVISAQPKMRSSQLLPAGRSLSLPSPSSSTASVLP 646
QY 904 PIRNSSPYSVIPQGMGNOGMINOGNLGNS-STGMIGNSASRPTMPSGEMAPQSSAVR 962
Db 647 PGLSLTITLAPPO-----DSDQCPSPDFGHDROLRLLSQPIQPMFMGSCDARQPSVVS 701
QY 963 VT-----CAATTSAMNRPVQGMIRNPASITMRRSSQOGORCTLOSO 1005
Db 702 RTGROVYVYAOQOVMPSPDSHPHTNSSASTPV--LIMQAAVLHPSFSPSPLO-----753
QY 1006 VMNIGSELEKMNMGPOYSQOQAPPNCTAP-----WPESTLPIQASFASONRQPFSS 1059
Db 754 -----PAAQOQPPPIYQAPRTSLHSEQPSDIL-----LSTFSQOQETTLGTA 794
QY 1060 PDDLCPHAAES 1072
Db 795 ATQSTPPQPPRPS 807

```

```

RESULT 14
US-08-971-188-8
; Sequence 8, Application US/08971188
; Patent No. 6326165
; GENERAL INFORMATION:
; APPLICANT: Wilson, Thomas G.
; APPLICANT: Heinrich, Julia N.
; TITLE OF INVENTION: RECOMBINANT BHLH-PAS/JHR POLYPEPTIDE AND ITS USE TO
; FILE REFERENCE: 082584/102
; CURRENT APPLICATION NUMBER: US/08/971.188
; EARLIER FILING DATE: 1997-11-17
; EARLIER APPLICATION NUMBER: 08/843.205
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 8
; LENGTH: 631
; TYPE: PRT
; ORGANISM: Drosophila
US-08-971-188-8

```

Query Match 3.8%; Score 286.5; DB 4; Length 631;

Best Local Similarity 22.6%; Pred. No. 2e-12; Mismatches 275; Indels 179; Gaps 35;

```

QY 7 NTSDPSRAETRRKKECPDQLGSPKRTKRNREQENKYYIELAELJFANFNIDIDNEFK 66
Db 5 NIDOKERFASRE-NHC-----EIERRRRKNKATAYITELSDMV-----PTGSALARK 49
QY 67 PDKCAIKETVYKQIRQIKDEKAAANIDEVQKSDVSSTG---OGVIDKDALGPMMLTAL 123

```

```

Db 50 PDKLILRMAYAHMKALR-----GTGNTSSDGTYYKPSFLTQDELKHLLEAA 96
QY 124 DGEFFVNVLE-GNVFVSENVTOYIRYNOEELMKSVSYLIHVGDTHEFVANKLPKSYVN 182
Db 97 DGEFFVNSCDGSRVLYVSDVTPVLYNTQSDWYGTSLYEHLHPDREKIRQLSTQESON 156
QY 183 GGS-----WSEGPFRNRNSHT-----FNCRMLVPLPDSSEGHNDNOAH-----222
Db 157 AGRILDLKSGTVYKKEGHOSMRLSMGARRGFTICARGVANVPESMVSGLHNLKQORNSLG 216
QY 223 -----KYETMOC-FAVSQPKS-----IKEEGEDLOS--CLICVAR-----RVPMKERP 262
Db 217 SRDGTNVAVVHCTGYIKWMPPTDMPRNHMERDMDMSHCICVLAIGRLQYTTAANDMS 276
QY 263 VLPSESTTTRODLOGKITSIDTSTMRAMKPGMEDLYRCLOKFAHQHEESYAKRH 322
Db 277 GSNNOSEFITHRAMCKTEFDORVNLIGYTPPELLKICYDFHR---EDOSHAKES 332
QY 323 HHEVLRO-GLAFSQIYRPSLDGT-LVAAGTKSKLIRSQTNNEPOLVTSIHLHREQMVC 380
Db 333 FDQVLYKQKQGFSLIYRARKANSEYVWLRTQAYAFNLPTDEVEYI-----VC 381
QY 381 VMNPDLTGQTM-GKPLNDISSNSP-----AHQALCSGNPGQDMLSSNINPEIN 428
Db 429 --RREL-TPVGSATNDGMYQTHMLAQAPPO-----QOQOQOQRPQSAQTTPVGYTY 478
QY 429 GPKEDMGKPRGFRGSG-GMNVHSGMQATTPOGNSVYALKMNSPSSSGMMPGQPTS-ML 486
Db 429 --RREL-TPVGSATNDGMYQTHMLAQAPPO-----QOQOQOQRPQSAQTTPVGYTY 478
QY 487 SPRHRMSPGAVG--SP--RIPPSQSPAGSLHSPVGVCSSTGNSHSYNSLNL-----537
Db 479 DTH--SPYSAGGTSPLAKITKSGTSP-----TYV-----APNSMALRPQOO 519
QY 538 ---QALSEGHGVSLGSSLASP-----DLKMN-----LQNSPVMMNPPLSKM 577
Db 520 QOQOQVTEGYOYQOTSPARSPSGPTYQLSAGNGRQOAPGAYQACP-----PPPNAP 575
QY 578 GSLSDKDCFGLXGPESECTGOAESCHPBGOKETNDP 615
Db 576 GMDWQOAGG-HRPHPTAHNRHHPAGGAGAP 612

```

```

RESULT 15
US-08-785-241-5
; Sequence 5, Application US/08785241
; Patent No. 5695963
; GENERAL INFORMATION:
; APPLICANT: McKnight, Steven L.
; APPLICANT: Russell, David W.
; TITLE OF INVENTION: Endothelial PAS Domain Protein
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
; STREET: 268 BUSH STREET, SUITE 3200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: 17-JAN-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: OSMAN, RICHARD A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: UTSD:1229

```



```

Db 128 QAGLTSP--NMGMGTSG--PNQGPQTQ--STGMNNSPVNOPAMGMNTG-----TNAGNPMGM 177
Oy 538 QALSGCHGV-----SLGSSLASPDLK-----MGNLONSVPVNMNPPPLS----- 575
Db 178 LAAGNGOGIMPNQVNGNSIGAGRGRODMQYRPNPBGMSAGNLLTEPLQOGSPQMGOTGLR 237
Oy 576 -----KMGSLDSKDCFGLYGEPSECTTGOAESCHPGEQKETN---DPNLPPAVSSERA- 626
Db 238 GPOPLKMGMMNPN---PYGSPYTQNPQOIGASGLGIQIKTYLSNNLSFPADKKAV 294
Oy 627 --DQSRLLHDSKGQTKLQLLTTKSDQMP---SPLASSLSTPNKSDTSGSLPGSGSTHGT 681
Db 295 PGGGMPNM---GQGPAPQV-----OQPGLVTPVAQ-----GMGSGAHT 329
Oy 682 SLKEKHILHRL-----QDSSSPVD-----LAKLTAEATGKD-- 714
Db 330 ADPEKRKLIOOVLVLLLAHAKCOREQAGEVROCNLPCHRTKMYLVNMTFCOSGSKSO 389
Oy 715 -----LSQESSSTAPGSEVTIKOEPVSPKKENALLRYL---LDKDDTKDIGLPE 761
Db 390 VAHCASSRQIISHMKNCIRHDCPVCLPLKNAGDKRNOPIITGATVYGLGNPSLSYG--QO 448
Oy 762 TTPKLERLDSKTDPAASNTKLA-----MKTE-----KEMSFEPGDPGS--ELD 804
Db 449 SAPNLSTV--SQIDPSSIERAYALGLPYOVNQMPQPOVQAKNQOQOGSPQGMRPMS 507
Oy 805 NLEETLDOLONS---QLQLEPDTTPGAPAGSVKQAIINDLMQITAEKSPVTPVGAQKT 861
Db 508 NMSASPMGVNGVGVQVTPSLSDSM-----LHSAINSONPMSENAVSPLGPMPT 558
Oy 862 AL-----RISOSTFNNPRGOLG----- 879
Db 559 AAQPSSTGIRKQMHEDITQDLRNHLVHKLVQAIPTTPPALKDRMENTVAYARKVED 618
Oy 880 -----RLPNONLEPLDITLQO-----PTGAGPFPPIRNSSPYVIP 915
Db 619 MYESANRNEYHYHLAEKIYKIOKELEEKRTRLQKQNNLPMAGGVPMVSMNPGMGP 678
Oy 916 OPGMNGNMGIMGNQNLGNSSTGMIGNSASRPTMPSGEMAPSSAVRYTCAATTSAMNRP 975
Db 679 QPGMTSN-----GLPDPDS--MIRGSVPNQMP----- 704
Oy 976 VOGMIRMPAASIPMRPSSQPCQROTLOSOVNMNIGPSELEMMGGPQYSSQOAPNQTAP 1035
Db 705 -----KITPOSGLNQFG-----QMSMAQPPVYPPROTTPPLQ--- 734
Oy 1036 WPESTILPIDOASFASQNRQPPGSSPDLLCPHPAAESPSEDEGALLDQVLYLALNFDGLEE 1095
Db 735 -----HHQQLAQPALNRP--MGYGRPMQOPSNQOQLPQ----- 767
Oy 1096 IDRALGIELVYSQOAVNPPEQESSQODSNIMLEOKAPVFFPOOYASQAOAMAQGS--YSPMQ 1152
Db 768 -----TOPPSGMMVNTNIP LAPSSGOAFVSQAOAMSSSCPVNSPIM 808
Oy 1153 DPNHETMGORPSYATLRMQPRGRLPRTGLVONQPNQNLRLQQLRHRLQAQONRQPLNNQJSN 1212
Db 809 PP-----GSQSHIHCPOLPOPALH-----QNSPSP----- 834
Oy 1213 VSNVNLTRPGVPTQAPINAGMLAOREILNQHRLROROMHQOQOVQORTLLMRGQGLNM 1272
Db 835 -----VPSKTP-----TPH-----HPPPSI-----GAQO 853
Oy 1273 TPSWAVBSGMPATMSNPRIPOANA-----QDFPPPNPTGISQODPFGTATTPOSPLMS 1327
Db 854 PPATTIPAPVPTPPAMPGPQSOALHPRPQRTPPTTQLPQOYVPSLPAPSSADPOQO 913
Oy 1328 PRMAHT-----QSPMMQSOANPAYOAPSDING 1355
Db 914 PRSQOSTAASVTPNAPILPPOPATPLSOPAVSIEG 949

```

```

PCT-US95-04682-2
: Sequence 2, Application PC/TUS9504682
: GENERAL INFORMATION:
: APPLICANT:
: TITLE OF INVENTION: NUCLEIC ACID ENCODING TRANSCRIPTIION
: TITLE OF INVENTION: FACTOR P300 AND USES OF P300
: NUMBER OF SEQUENCES: 13
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Weingarten, Schurgin, Gagnebin & Hayes
: STREET: Ten Post Office Square
: CITY: Boston
: STATE: MA
: COUNTRY: US
: ZIP: 02109
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US95/04682
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/227,536
: FILING DATE: 14-April-1994
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: Holliday C. Heine, Ph.D.
: REGISTRATION NUMBER: 34,346
: REFERENCE/DOCKET NUMBER: DPCI-308Xq999
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (617) 542-2290
: TELEFAX: (617) 451-0313
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2414 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: PCT-US95-04682-2

Query Match 3 6%; Score 277; DB 5; Length 2414;
Best Local Similarity 18.9%; Pred. No. 8.2e-11;
Matches 222; Conservative 117; Mismatches 363; Indels 474; Gaps 50;

Oy 381 VANNPDLTGOTMCKPLNPI-----SSNSPAHQALCSGNPGQDMLTSSNINFPING 429
Db 47 LINSTELGITNGCDINOQLQTSIGWYODASAKHIKOLSELLRSGS-----SPNLNMGVCG 99
Oy 430 PREQMGMPGRFSGSGGMNHSVGMQATTPQGSNVYALKMNSPSSQSPGM-----NPGQ 481
Db 100 PQOVNA-----SOAQOSSPGLGLINSKVPKWT 127
Oy 482 PTSMLSPRHRMSPVAGSPRIPPSQSPAGSLHSPV---GYCSSNGNSHSYTNSSLN-AL 537
Db 128 QAGLTSP--NMGMGTSG--PNQGPQTQ--STGMNNSPVNOPAMGMNTG-----TNAGNPMGM 177
Oy 538 QALSGCHGV-----SLGSSLASPDLK-----MGNLONSVPVNMNPPPLS----- 575
Db 178 LAAGNGOGIMPNQVNGNSIGAGRGRODMQYRPNPBGMSAGNLLTEPLQOGSPQMGOTGLR 237
Oy 576 -----KMGSLDSKDCFGLYGEPSECTTGOAESCHPGEQKETN---DPNLPPAVSSERA- 626
Db 238 GPOPLKMGMMNPN---PYGSPYTQNPQOIGASGLGIQIKTYLSNNLSFPADKKAV 294
Oy 627 --DQSRLLHDSKGQTKLQLLTTKSDQMP---SPLASSLSTPNKSDTSGSLPGSGSTHGT 681
Db 295 PGGGMPNM---GQGPAPQV-----OQPGLVTPVAQ-----GMGSGAHT 329
Oy 682 SLKEKHILHRL-----QDSSSPVD-----LAKLTAEATGKD-- 714

```

```

Db 330 ADEBRKLLIOOLVLLLHAKCORREOANGVEBOCNLPHCRITKNVNLNHTHCOSGKSCQ 389
QY 715 -----LSQESSSTAGSEVITIKOEVPSPKKNENALLRYL---LDKDDTKDIGLPE 761
Db 390 VAHCASSROIISHMKNTKTRDCEVCLPLKNAQKRNQOPLITGAPVGLSPSSLGAV-QQ 448
QY 762 ITPKLERLDSKTPDASNTKLLIA-----MKE-----KEEMSEFGDQGS--ELD 804
Db 449 SARNLSTV-SQIDPSSITERAYALGLPYQVQNPOTPOVQAKNOQNOQPOGSPQGMKPMPS 507
QY 805 NLEBIIIDDLONS---QLPOLFPTTRPGAPAGSDKAIINDLMQITRENPPVTPVYQKRT 861
Db 508 NMSASPMGVGVGVQVTPSLSDSM-----LHANSQNPMMSENASVSLGMP 558
QY 862 AL-----RISQSTFNNRPGOLG-----879
Db 559 AARPSTTGIRKQNHEDITQOLRNHLVHKLVOALFPTTDPALALDRMENLVAAKRYEGD 618
QY 880 -----RLPNQNLPLDITLOS-----PTGAPRPPIRNSPYSVIP 915
Db 619 MYESANRAEYVHLLAKIKYIKQKLEEKRRTRLOKONMLPNAAGVPMNPGPMNGQP 678
QY 916 QPGMKGNOGIGNOGNLNGNSTGMIGNSASRPTMPSEMAPOSSAVYTCATTSANMRP 975
Db 679 QPGMTSN-----GRLPDS--MIRGSVPNQMP-----704
QY 976 VQGMGINPAAISIPMRSSQPGOROTLOSQVMNIGPSELEMMNGGPQYSSQAAPPNOTAP 1035
Db 705 -----RITPQSLNPG-----QMSAAPRTVIRQTPPLQ---734
QY 1036 WPESILPIDQASFRASQNRQPGSSPDDLCPHRAESPSDEGALLDOLYALRNFDGLEE 1095
Db 735 -----HHQOLAQAGALNPP--MGYGPMDQPSNQGOFLLPQ-----767
QY 1096 IDRALGIPELVQSQAADVPQSSQDSNIMLEQKAPVFPQQAYSQAQMAQS---VSPMQ 1152
Db 768 -----TOPPSQGMNVTNIPLABSSGAPVSAQMSSSSCPVSPIM 808
QY 1153 DPNEHTMGQRPYSATLMQRPGLRPTGLVQONQPNQRLQLOHRLQAQNRQPLMNQISN 1212
Db 809 PR-----GSGGSHHCQQLQRPALH-----QNSPSP-----834
QY 1213 VSNVNTLIRGCVTPQAFINQMILAQORREILNHLRORHMOOQVOQRTLMRGGLM 1272
Db 835 -----VPSKRP-----TPH-----HTPSSI-----GAQ 853
QY 1273 TPSSVAPSGMPATMSNRIPOANA-----QOPFPPNYGISQODPDGFGATTPQSPILMS 1327
Db 854 PRATTTPAVPTPPAMPBGFSQALHPPRQOTPPPTTQLPQOVQPSLPAPASADOPQOQ 913
QY 1328 PRMAHT-----QSPMQOSQANPAYQAPSDING 1355
Db 914 PRSQOSTAASVPTPPMAPLLPPQATPLSQPAVSIEG 949

```

```

RESULT 18
US-08-785-241-6
; Sequence 6, Application us/08785241
; Patent No. 5695963
; GENERAL INFORMATION:
; APPLICANT: McKnight, Steven L.
; APPLICANT: Tian, Hu1
; APPLICANT: Russell, David W.
; TITLE OF INVENTION: Endothelial PAS Domain Protein
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
; STREET: 268 BUSH STREET, SUITE 3200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:

```

```

; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/785,241
; FILING DATE: 17-JAN-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: OSMAN, RICHARD A.
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: UTSD:1229
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 343-4341
; TELEFAX: (415) 343-4342
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 826 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-785-241-6

```

```

Query Match 3.5%; Score 269.5; DB 1; Length 826;
Best Local Similarity 19.3%; Pred. No. 5,2e-11;
Matches 186; Conservative 150; Mismatches 337; Indels 289; Gaps 41;

```

```

QY 1 MSGKENTSDPSRAETRRKRECPDOLGSPKRNTEKRNKEOEKYLELAELIFANFNDI 60
Db 1 MEGAG-GANDKKRISSEKRE-----KSRDAASRRKRESEVEYLAHQPLRHNS 51
QY 61 DNFENPKDCAIKETVYKQIRQIKEDQKAAANIDEVOKSDVSTGQYIDKALGPMML 120
Db 52 SHL-----DKASVRLITSTYR-VKRLDAGDDLIEDDKRAQNM-----FYL 93
QY 121 EALDGEFFVNLGVNVESENVTQYLRNQEELMKNKSVYSLHVDHTEFVNLLPKST 180
Db 94 KALDGEFVMTLDDGDMYISDNVKNKMGTLQFELTGHVDFPHRCDHEEMRLTHR-- 151
QY 181 VNGSMSGEPFRKNSHTFNCRMLVYKRLPSEEGHNOQENHOKYETMOCRAVSQPSIKE 240
Db 152 -NG-----LVK-----KKEQONTQSFELMKCTLTSGRTMT 184
QY 241 EG-----EDLQSLICVARVPMKERVLP--SSESFTT 272
Db 185 KSAFTKVLHCTGHIHYDTNSNQPGCYKKRPMTCLVLICERLPHNSNIEPLDSTFELS 244
QY 273 RQDLQKITSLDYS-TWRAAMKPGMEDLYRRCIOKFAHQHESVSYAKRHHHEVLROGL 331
Db 245 RHSLDMKFYSYCDERITELMGEYR--EELLGRSIYEY--HALDS-DHLTKTHHDMFTKQ 299
QY 332 AFQOIYRFSLSDGTIVAQAQKSKLISQTNPEQOLVLSLH-----MLHRCQNV 380
Db 300 VTTGQYRMLAKRGGYVWVEVQATVYNTKNSQPCQICVAVYVSGIIGHDLISLQDTEC 359
QY 381 VVNP-DLTGQTMKPLNPISNSPAH-----QALCSGNGQDMTSSNINFPINGP 430
Db 360 VLKRVESSDMKMQLFTKYVSESDTSFLDKLKEKPEALITLLAAGDTIIS-IDF-----413
QY 431 KEQGMGMPGRFGSGGNHVSQMAQATTPOGSNYALKMNSPQSSPGMNPQOPTSMLSPRH 490
Db 414 -----GSDNTEPDQOLEVPLVYN--DVMLSPNE 441
QY 491 RM-SPGVAGSPRIIPPSFAGSLHSPVGVCSSTGNSHSTNSLNAQLAS--EGHCYSL 548
Db 442 KIONINLAMP-----LFTAETPKPL-----RSSADPALNOEVALKLEPPESL 485
QY 549 GSSIASPDLKMGNLONSVPVMMNPPLSKMGLSDKDCFGLYGPESEGTGOAESCHPGE 608
Db 486 ELSTFMQIO-----DQTP-----SPSDGSTNO--SSPEPNS 515

```



```

CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 4225 Executive Square, Suite 1400
CITY: La Jolla
STATE: CA
COUNTRY: USA
ZIP: 92037

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/915,213
FILING DATE: 20-AUG-1997
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/480,473
FILING DATE: 06-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Halle, Lisa A.
REGISTRATION NUMBER: 38,347
REFERENCE/DOCKET NUMBER: 07265/053001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619/678-5070
TELEFAX: 619/678-5099
INFORMATION FOR SEQ. ID NO. 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 826 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-915-213-2

```

```

Query Match          3.5%; Score 269.5; DB 3; Length 826;
Best Local Similarity 19.3%; Pred. No. 5,2e-11;
Matches 186; Conservative 150; Mismatches 337; Indels 289; Gaps 41;

```

```

QY 1 MSGGENTSPSRAETRKRECPDQIGSPKRNTERKRNQENKYEIDELAFANFNDI 60
DB 1 MEGAG-GANDKKKISSRRKE-----KSRDAKRRKSESEVEYELAHQPLPHNVS 51
QY 61 DNFNFKPKCALIKETVKQIRKOEKAAANIDEVOKSDVSTGGVIDKDALGPMML 120
DB 52 SHL-----DKASVRLITSTYLR-VKRLLDAGDLIEDMKAKQMC-----FYL 93
QY 121 EALDGEFFVNLGNVVFSENYTOYLRYNOBELMKNVSYLAVGDHTEFVKNLPRST 180
DB 94 KALDGFVNLTDGDMITISDNVKNYMGLTQFELTGHVSVDFTGHPCHHEMRMLTHR-- 151
QY 181 VNGGSMGSEPPRRNSHTFNCRLVPLPDEEGHNDQEAHQKYEITMOCFAVSQPSIKE 240
DB 152 -NG-----LVK-----KGEQNTORSFPLMKCTLTLSRGRTMNT 184
QY 241 EG-----BDLQSLICVARVPKMERPLP--SSESFTT 272
DB 185 KSAFWKVLHCTGHIVDTNSNQPGYKKRPMTCLVLICEPIHPNSINELPDSKFTLS 244
QY 273 RODLOGKITSLDTS--TYRAAMKPEMEDLVRCIOKFAHQEGESVSYAKRHHHEVLROGL 331
DB 245 RHSIDMKFSYCDERITELMGEYR--EELIGRSIYEY--HALDS-DHITKTHDMFTKGO 299
QY 332 AFSQIYFSLSDGTIVAAQKRSKLIRSQITNEPQVLISLH-----MLHREQNVC 380
DB 300 VTTGVGVMLAKRGQYVVEVQATVYINTKNSQPCQICVAVVSGIIOHDLITSLQOTEC 359
QY 381 VAMP-DLTGOTMGKPLNPISNSPAH-----QALCSGNPGODMTLSSNINEPINGP 430
DB 360 VLKPVESDDMKQTLFTKVSSEDTSSLEFDLKKPEKPLITLLAPAGTTIIS-LDF----- 413
QY 431 KEQMGMPMGRRFGSGGNHVSIGMQATTPGGSNVALKKNSSQSSPCGNPQOPTSMLSPRH 490

```

```

DB 414 -----GSNDTEFDQQLFEVPLYN---DVMLESPNE 441
QY 491 RM-SPVAGSPRIPQSPAGSLHSPVGCSTGNSHTYNSLNAOLMS'-EGHGVSL 548
DB 442 KIONINLAMP-----LPTAETPKPL-----RSSADPALNQEVALKIEPPESL 485
QY 549 GSSLASPDLMKGNLONSPVNMNPPILSKMGSLDKDCFGLYGSPSECTGQAESSCHPE 608
DB 486 ELFTMPQIQ-----DQTP-----SPDSGSTRQ--SSPEPNS 515
QY 609 QKE-----INDNLPRAVSSERADQSRHDSKGQTKL-LQLLT-----T 647
DB 516 PSEYCFYVDSDVNEKLELVEKFAEDTEAKNPFSTQDTDLLEMALRYIPMDDDFOLR 575
QY 648 KSDMEPSPLASLSDTNKDSGSL-----PGSGSTGTSLEKXKILH----- 691
DB 576 SFDQL--SPLESSASAPESAPOSTYVTVQOQIOEPTANATTTTATTDLAKTVTDRE 633
QY 692 --RLDSSSPVDLAKITAEAT-----GKDLQSSSTAPGS----- 726
DB 634 DIKLASPEPTHIHKETTTSATSPYRDQSRTPASPNRAGKVIEQTEKSHRSPVLSV 693
QY 727 ----EVTIKQEPVSPK--KKNALLRYLLDKDT--KDIGLEPITTKLERLDSKTPASN 778
DB 694 ALSQRTVPEEELNPKILALONQRRKMHDSLFQAVGIGTL-----LQPPDHAAT 747
QY 779 TKLIAMKTEKEEMSFEPDQSGSELDNLEIILDDL-----QNSQLPOLFP-DTRPG 828
DB 748 TSL-----SKRVKCGKSSSEONGMEOKTITILPDLACRLIGQSMDESGLPOLTSYCEVN 803
QY 829 AP 830
DB 804 AP 805

```

```

RESULT 21
US-09-148-547-2
; Sequence 2, Application US/09148547
; Patent No. 6124131
; GENERAL INFORMATION:
; APPLICANT: Semenza, Gregg L.
; TITLE OF INVENTION: Hypoxia Inducible Factor-1 and Methods of Use
; FILE REFERENCE: 07265/151001
; CURRENT APPLICATION NUMBER: US/09/148,547
; NUMBER OF SEQ. ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.0
; SEQ. ID NO. 2
; LENGTH: 826
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-148-547-2

```

```

Query Match          3.5%; Score 269.5; DB 3; Length 826;
Best Local Similarity 19.3%; Pred. No. 5,2e-11;
Matches 186; Conservative 150; Mismatches 337; Indels 289; Gaps 41;

```

```

QY 1 MSGGENTSPSRAETRKRECPDQIGSPKRNTERKRNQENKYEIDELAFANFNDI 60
DB 1 MEGAG-GANDKKKISSRRKE-----KSRDAKRRKSESEVEYELAHQPLPHNVS 51
QY 61 DNFNFKPKCALIKETVKQIRKOEKAAANIDEVOKSDVSTGGVIDKDALGPMML 120
DB 52 SHL-----DKASVRLITSTYLR-VKRLLDAGDLIEDMKAKQMC-----FYL 93
QY 121 EALDGEFFVNLGNVVFSENYTOYLRYNOBELMKNVSYLAVGDHTEFVKNLPRST 180
DB 94 KALDGFVNLTDGDMITISDNVKNYMGLTQFELTGHVSVDFTGHPCHHEMRMLTHR-- 151
QY 181 VNGGSMGSEPPRRNSHTFNCRLVPLPDEEGHNDQEAHQKYEITMOCFAVSQPSIKE 240
DB 152 -NG-----LVK-----KGEQNTORSFPLMKCTLTLSRGRTMNT 184

```

```

OY 241 EG-----EDLOCLICVARRVPMKRPVLP-SSESTTT 272
DB 185 KSAWKVLHCTGHIHVYDTSNQPOCGYKKRPMTCVLICERIPRPSNIEIPLDKTFLS 244
OY 273 RODLOGKITSLDTS--TMRAMKPGMEDLVRRICIOFHAOHEGESVYAKRHHNEVLRGOL 331
DB 245 RHLIDMKFSYCDERTELMGYER--EELLGRSIEY--HALDS-DHLTKTHHDMFTKGO 299
OY 332 AFSQIYREFSLDGLVAAQTKSLIRSQTNPEOLVISLH-----MLHREDQVNC 380
DB 300 VTTGQYRMLAKRGYVWVETQATVIYNTKNSOPQICVNVVSGIIQHDILFSLQOTEC 359
OY 381 VMNP-DLTGOTMGKPLNFISSNSPAH-----QALCSNPGODMTLSSNINPINGP 430
DB 360 VLKPESSDMKMTQLFTEVSEEDTSSLFDKLLKEPDALTLLAPAGDTIIS-LDF----- 413
OY 431 KEQMGMPGRFGSGGGMNHNVSQMOATTPQGSNYALKNMSPSQSPGMNGOPTSMLSPRH 490
DB 414 -----GSDNETDDQOLEEYPLVN---DVMLPSPNE 441
OY 491 RM-SPGVAGSPRIPPSQSPAGSLSPYGVCSSTGNSSHYSNLSMALQALS-BEGHVSIL 548
DB 442 KLQMINILAMSP-----LPTAETPKPL-----RSSADPALNOEVALKLEPNPESL 485
OY 549 GSSLASPDLKMGNIQNSPVNNNPPPLSKMGLSKDCGFLGEPSEGTGGAESSCHPGE 608
DB 486 ELSTFMQIO-----DQTP-----SPDSGSTRQ--SSPEPNS 515
OY 609 QKE-----TNDPMLPAVSSERADGOSRLHDSKGOTKL-LQLLT-----T 647
DB 516 PSECFYVDSDMVNEFEKLEVEKLEAEDTEAKNPFSTODTDLLEMLAPYIPMDDDFOLR 575
OY 648 KSDMEPSPLASSISDTRKDKSTGSL-----PGSGTHGSLKEKKILH----- 691
DB 576 SFDOL--SPLSSSSASPSASPOSTVTVYQOTOIOPETANMTTATTDKLTATYKDRME 633
OY 692 --RLLODSSFPVLDLAKLTAET-----GRDLSQESSSTAPGS----- 726
DB 634 DIKLILASPSPTHIHKEETTSATSSPYRDTQSRASPNRAGGVIEQTEKSPRSNVLVS 693
OY 727 ----EVTIKQEPVSPK--KRENALIRYLKDDOT--KDIGLEPTPKLERLDSKTDPSAN 778
DB 694 ALSORTVPEELNPKILALQNAQRKRKMEHDSLFQAVGIGTL-----LQOPDDHAAT 747
OY 779 TKLIAMTEKEMSPEDQPSGSELDNLEITLDL-----QNSOLPOLFP-DTRPG 828
DB 748 TSL-----SMKRVKCGCKSSBQMGEMOKTIIILPSDLACRLGOSMDESGLPOLTYDCEVN 803
OY 829 AP 830
DB 804 AP 805

```

```

RESULT 22
US-09-235-217-2
: Sequence 2, Application us/09235217
: Patent No. 6222018
: GENERAL INFORMATION:
: APPLICANT: Semenza, Gregg L.
: TITLE OF INVENTION: HYPOXIA INDUCIBLE FACTOR-1 AND METHOD OF USE
: NUMBER OF SEQUENCES: 64
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Fish & Richardson P.C.
: STREET: 4225 Executive Square, Suite 1400
: CITY: La Jolla
: STATE: CA
: COUNTRY: USA
: ZIP: 92037
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS

```

```

: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/235,217
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/480,473
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Haile, Lisa A.
: REGISTRATION NUMBER: 38,347
: REFERENCE/DOCKET NUMBER: 07265/053001
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 619/678-5070
: TELEFAX: 619/678-5099
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 826 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
US-09-235-217-2

```

```

Query Match 3.5%; Score 269.5; DB 4; Length 826;
Best Local Similarity 19.3%; Pred. No. 5.2e-11;
Matches 186; Conservative 150; Mismatches 337; Indels 289; Gaps 41;

```

```

OY 1 MSGGENTSPPSRATRRKRCPPDGLSPKRRTEKRRNEQENKYEIELAELIFANFNDI 60
DB 1 MEGAG-GANDKKKISERRKE-----KSRDAARSRSKSESEFYELANQLPLPHNVS 51
OY 61 DNFNFKPKCALIKETVQIQOIKOEKAAANANIDEVQKSDVSSGTGCVIDKDALCPMML 120
DB 52 SHL-----DKASVMRLTISYLR-VKRLDAGDLDIEDDKAKQAMNC-----FYL 93
OY 121 EALDFFEFVNLQGVNVSENVTOYLRYNOEELMKNKSVYGIILVGHDTFEVKNILPKSI 180
DB 94 KALDGFVAVVLDLDDGMITISDNVKNKYKGLTPEELGHSVFEPFTIPRCHBEKREMLTIR- 151
OY 181 VNGSWSGEPFRNSHTFNCMLVYKPLPDSSEBHDQEAHOKYETMOCFAVSOPKSIKE 240
DB 152 -NG-----LVK-----KKEQNTQSFPLRMKCTLTSRGRTWNI 184
OY 241 EG-----EDLOCLICVARRVPMKRPVLP-SSESTTT 272
DB 185 KSAWKVLHCTGHIHVYDTSNQPOCGYKKRPMTCVLICERIPRPSNIEIPLDKTFLS 244
OY 273 RODLOGKITSLDTS--TMRAMKPGMEDLVRRICIOFHAOHEGESVYAKRHHNEVLRGOL 331
DB 245 RHLIDMKFSYCDERTELMGYER--EELLGRSIEY--HALDS-DHLTKTHHDMFTKGO 299
OY 332 AFSQIYREFSLDGLVAAQTKSLIRSQTNPEOLVISLH-----MLHREDQVNC 380
DB 300 VTTGQYRMLAKRGYVWVETQATVIYNTKNSOPQICVNVVSGIIQHDILFSLQOTEC 359
OY 381 VMNP-DLTGOTMGKPLNFISSNSPAH-----QALCSNPGODMTLSSNINPINGP 430
DB 360 VLKPESSDMKMTQLFTEVSEEDTSSLFDKLLKEPDALTLLAPAGDTIIS-LDF----- 413
OY 431 KEQMGMPGRFGSGGGMNHNVSQMOATTPQGSNYALKNMSPSQSPGMNGOPTSMLSPRH 490
DB 414 -----GSDNETDDQOLEEYPLVN---DVMLPSPNE 441
OY 491 RM-SPGVAGSPRIPPSQSPAGSLSPYGVCSSTGNSSHYSNLSMALQALS-BEGHVSIL 548
DB 442 KLQMINILAMSP-----LPTAETPKPL-----RSSADPALNOEVALKLEPNPESL 485
OY 549 GSSLASPDLKMGNIQNSPVNNNPPPLSKMGLSKDCGFLGEPSEGTGGAESSCHPGE 608
DB 486 ELSTFMQIO-----DQTP-----SPDSGSTRQ--SSPEPNS 515
OY 609 QKE-----TNDPMLPAVSSERADGOSRLHDSKGOTKL-LQLLT-----T 647

```

Db 516 PSRYCYVSDWVNEFKLELVEKLFADDEAKNPFSTODDLDLEMLAPYIPMDDDFOLR 575  
 QY 648 KSDOMEPPPLASLSDTNKDSGSL-----PGSGTHGTSLKXKHLH----- 691  
 Db 576 SFQDL--SPLESSASBPESAPOSTYVFOQTQIDEPFANATTTATTDLKLVTYDKRME 633  
 QY 692 --RLDODSSPVDLAKLTAEAT-----GKLDOSBSSSTAPGS----- 726  
 Db 634 DIKILASPSPTIHKETTSATSSPYRDTQSRTPASPNRAGKGVLEQTEKSHRSPNVLAV 693  
 QY 727 ----EVTIKOEPVSPK--KENALLRYLDDKDT--KDIGLPEITPKLERLDSKTPASN 778  
 Db 694 ALSQRTTVEEELNPKILALQNAQRKRMHDSGLFQAVGIGTL-----LQPDHNAAT 747  
 QY 779 TKLIAMKTEKESFEPGDPGSELNLEILDDL-----QNSQLPOLFP--DTRPG 828  
 Db 748 TSL-----SWKRVGCKSSQNGMEOKTIIILIPSLDLACRLLGSMDSGSLPOLSTYDCEVN 803  
 QY 829 AP 830  
 Db 804 AP 805

RESULT 23  
 PCT-US96-10251-2  
 ; Sequence 2, Application PC/TUS9610251  
 ; GENERAL INFORMATION:  
 ; APPLICANT: The Johns Hopkins University School of Medicine  
 ; TITLE OF INVENTION: HYPOXIA INDUCIBLE FACTOR-1 AND METHOD OF USE  
 ; NUMBER OF SEQUENCES: 35  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Fish & Richardson P.C.  
 ; STREET: 4225 Executive Square, Suite 1400  
 ; CITY: La Jolla  
 ; STATE: CA  
 ; COUNTRY: USA  
 ; ZIP: 92037  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: floppy disk  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: PCT/US96/10251  
 ; FILING DATE: 06-JUN-1996  
 ; CLASSIFICATION:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Haile, Lisa A.  
 ; REGISTRATION NUMBER: 38,347  
 ; REFERENCE/DOCKET NUMBER: 07265/053WO1  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 619/678-5070  
 ; TELEFAX: 619/678-5099  
 ; INFORMATION FOR SEQ ID NO: 2:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 826 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; PCT-US96-10251-2

Query Match 3.5%; Score 269.5; DB 5; Length 826;  
 Best local similarity 19.3%; Pred. No. 5, 2e-11;  
 Matches 186; Conservative 150; Mismatches 337; Indels 289; Gaps 41;

QY 1 MSGMGENTSDPSAETRRKRRCPDGLGSPKRTNTERKRNREQNKYILEELAFIFANFNDI 60  
 Db 1 MEGAG--GANDKKKISSRKE-----KSRDAARSRGRKESSEVEYELAHQPLPHNVS 51  
 QY 61 DNENFRPKAILKEVYKQIRQIKDEKAAANIDEVQKSDVSTGGGVLDKALGPMML 120

Db 52 SHL-----DRASYWRLTISYLR--VRKLLDAGDLDIEDMKQAQNC-----EYL 93  
 QY 121 EALDGEFFVYVNEGANVFSENVYQYLRNOEELMKRSYVSLIHVGDHTEFYKNLPKSI 180  
 Db 94 KALDGFVWVLTDDGDMITISYDNVKNKMGTLQFELTHSHSFDTTHPDHREMKMLHR-- 151  
 QY 181 VNGSNWSEPPRRNSHTFNCRLVKPLPDSEEGHNDONAHOKYETMOCFAVSQPKSKE 240  
 Db 152 -NG-----LVK-----KGEQNTQORSFLELMKCTLSRGRTNMI 184  
 QY 241 EG-----EDQSLICARVYKMERVLV--SSSEFTT 272  
 Db 185 KSATWKLVLCTGHIVYDTNSQPOCGYKRPWTCLVLCEPRLPHNSIEIRLDSKTEFLS 244  
 QY 273 RODLGKITSLDTS--TMRAMKPGMEDLVRCIOKFEHNAOHESESAYAKNNHNEVLQGL 331  
 Db 245 RSLDMKFSYCDERITELMGYER--BELIGRSLYEY--HALDS--DHLKTHNDMTKQ 299  
 QY 332 AFSQIYRFSLSDDTLVAOATKSKLIRSQTTNEPOLYISLH-----MLHREQNV 380  
 Db 300 VTTGQYRMLAKRGYVWVEVTOATVIYNTKNSOPQOCIVCVNVVSGIIQHDLIFFSLDQTEC 359  
 QY 381 YANP--DLTGOTMGRPLNPTSSNSPAH-----QALCSNPGQDMTSSNINPFINP 430  
 Db 360 VLKPVESDMKMTQLETKVESEDTSLFPLKKEPDALTLAAGADTITS--LDF----- 413  
 QY 431 KEQMGPMRFGSGGGMNVHSGQATTPQGSNVYALKMNSPQSPQMPNPGQPTSMLSRN 490  
 Db 414 -----GSDNTEHDQGLEVPRLXN--DYMLSPNVE 441  
 QY 491 RM-SPEGVAGSPRIIPSOFPAGSLSPGVGVCSTGNSHTYSSNLALDLS--EGHGVSL 548  
 Db 442 KIONINLAMP-----LPTAETPKPL-----RSSADPALNQEVALKLEPNESL 485  
 QY 549 GSSLASPDLMKMLQNSPVNMPNPPRLSKMSLDSKDCFGLYGEPSGCTGOAESCHPE 608  
 Db 486 ELSEFTMPQIQ-----DQTP-----SPDGSSTRQ--SSPEPNS 515  
 QY 609 QKE-----TNDPLPRAVSEERADGQSRHLSKGOTKL-LQLLT-----T 647  
 Db 516 PSRYCYVSDWVNEFKLELVEKLFADDEAKNPFSTODDLDLEMLAPYIPMDDDFOLR 575  
 QY 648 KSDOMEPPPLASLSDTNKDSGSL-----PGSGTHGTSLKXKHLH----- 691  
 Db 576 SFQDL--SPLESSASBPESAPOSTYVFOQTQIDEPFANATTTATTDLKLVTYDKRME 633  
 QY 692 --RLDODSSPVDLAKLTAEAT-----GKLDOSBSSSTAPGS----- 726  
 Db 634 DIKILASPSPTIHKETTSATSSPYRDTQSRTPASPNRAGKGVLEQTEKSHRSPNVLAV 693  
 QY 727 ----EVTIKOEPVSPK--KENALLRYLDDKDT--KDIGLPEITPKLERLDSKTPASN 778  
 Db 694 ALSQRTTVEEELNPKILALQNAQRKRMHDSGLFQAVGIGTL-----LQPDHNAAT 747  
 QY 779 TKLIAMKTEKESFEPGDPGSELNLEILDDL-----QNSQLPOLFP--DTRPG 828  
 Db 748 TSL-----SWKRVGCKSSQNGMEOKTIIILIPSLDLACRLLGSMDSGSLPOLSTYDCEVN 803  
 QY 829 AP 830  
 Db 804 AP 805

RESULT 24  
 US-08-971-188-10  
 ; Sequence 10, Application US/08971188  
 ; Patent No. 6326165  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Wilson, Thomas G.  
 ; APPLICANT: Heinrich, Julia N.  
 ; TITLE OF INVENTION: RECOMBINANT BHLH-PAS/JHR POLYPEPTIDE AND ITS USE TO  
 ; FILE REFERENCE: 082584/102



```

? CURRENT APPLICATION NUMBER: US-08/971,188
? CURRENT FILING DATE: 1997-11-17
? EARLIER APPLICATION NUMBER: 08/843,205
? EARLIER FILING DATE: 1997-04-14
? NUMBER OF SEQ ID NOS: 14
? SOFTWARE: Patentin Ver. 2.0
? SEQ ID NO 10
? LENGTH: 626
? TYPE: PRT
? ORGANISM: brain and muscle ARNT-like protein a
? OS-08-971-188-10

```

Query Match	3.5%	Score 269	DB 4	Length 626
Best Local Similarity	24.3%	Pred. No. 3.6e-11		
Matches 142	Conservative 82	Mismatches 230	Indels 130	Gaps 27

```

OY 35 EKRRREONKYEELAEELIFANFNNDIDFNFRKPOCALKEVNOIROIK-EQEKAAAN 93
    ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 81 EKRRRDKNNSIDELASLV-----PTCNMSKRLLDLVLRMAVOMHRLRGATNPYTEAN 136
OY 94 IDEQKSDVSSTGGGIDKDALGPMLEALDGFEEVVMLE-GNVFVSENVTOYLRYNOE 152
    || : : : : : || : : : : : || : : : : : || : : : : : || : : : : :
Db 137 YKPYFLSD-----DELKHLILRAAGFLFVVGCDGSKILFVSESVFKILNIYSON 185
OY 153 ELANKSVYSLIHVGDIHEEVKNLLPKS-----IVNGSWSGEPBR 192
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 186 DLIGOSLFEDYLAHPDIAK-YKEQSLSSDPTADPRERLIDAKTGLPVKTDITDIPGSRILCSGAR 244
OY 193 RNSHTFNCRLV-KPLDSEEEGHNDQEAHOKYETMOCFANVSQPSIR----- 239
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 245 R---SFECKRMKCNBPVKVEDKDPSTCSKKAKRKSPTCTHSTGCLSWPRTKMGLED 301
OY 240 -----EEEDLQSLICVARR-----VPMK-EKRVLPSESEFTTRDLOGKITSLD-TSTM 289
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 302 NEPDNECNL-SCVLVAGRLSHVYPPQVNEIKVSMELYSRNAIDGRFVVDORALAI 360
OY 290 AAMKPGMEDLVRCIQKFNHOGESVYAKRHHEVL-ROGLAFSOLITRELSLDTGLV 347
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 361 LAYLP-CELLGTSCYEYFHQ---DIGHLEACHROVLQDTREKIT-TNCYCFKIKDGSFI 414
OY 348 AAOIKSLINSQATNEPQVLISLMLHREQVVCVMPNPDLTGQTMKRLPILSSNSPADAQ 407
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 415 TLKSRMFSFMPKVEEYIYS-----TIVVLANYLEG---GDPFFPOLTAAP----- 460
OY 408 ICSGNPQDMTLSSINIFPIGPKRQMGMPRGSGSGMHNVSQ---MOATTPGGSNYA 464
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 461 -----HMDMSLPGEGEGPKTHPTVVGIPGCTAGACKIGRMALAEIMEIHRIGS--- 512
OY 465 LKMNPSQSSFGMNPQOPTSMLSPRHNSPG-----VAGSPRIPEQSPSPAGSLHSPGV 519
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 513 -----SPSSCGSSPLNITS-TPPDASSPGKKILMNGTGPDISSGLSGQAQENP--- 562
OY 520 CTSSTGNSHSTNSSLNALQALSEGHVSL-----GSSLASPD 566
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 563 -----GYPVSDS-----SILGENPHIGIDIMDQGGSSPSND 596

```

RESULT 25  
US-08-816-693A-51  
: Sequence 51, Application US/0881693A  
: Patent No. 5874241  
: GENERAL INFORMATION:  
: APPLICANT: Takahashi, Joseph S  
: APPLICANT: Turek, Fred W  
: APPLICANT: Pinto, Lawrence H  
: TITLE OF INVENTION: Clock Gene and Gene Product  
: NUMBER OF SEQUENCES: 53  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: Dressler, Rocky, Minnow & Katz  
: STREET: Two Prudential Plaza, Suite 4700  
: CITY: Chicago  
: STATE: Illinois

```

1      COUNTRY: USA
2      ZIP: 60601
3
4      COMPUTER READABLE FORM:
5
6      MEDIUM TYPE: Floppy disk
7      COMPUTER: IBM PC compatible
8      OPERATING SYSTEM: PC-DOS/MS-DOS
9      SOFTWARE: Patentin Release #1.0, Version #1.3GD
10
11     CURRENT APPLICATION DATA:
12
13     APPLICATION NUMBER: US/08/816,693A
14
15     FILING DATE:
16
17     CLASSIFICATION: 435
18
19     ATTORNEY/AGENT INFORMATION:
20
21     NAME: NO. 5874241thrup, Thomas E
22
23     REGISTRATION NUMBER: 33,268
24
25     TELECOMMUNICATION INFORMATION:
26
27     TELEPHONE: 312-616-5400
28     TELEFAX: 312-616-5460
29
30     INFORMATION FOR SEQ ID NO: 51:
31
32     SEQUENCE CHARACTERISTICS:
33
34     LENGTH: 747 amino acids
35     TYPE: amino acid
36
37     STRANDEDNESS: single
38
39     TOPOLOGY: Linear
40
41     MOLECULE TYPE: peptide
42
43     OS-08-816-693A-51

```

Query Match	3.58;	Score 267;	DB 2;	Length 747;
Best Local Similarity	20.98;	Pred. NO. 6.7e-11;		
Matches 201;	Conservative 133;	Mismatches 345;	Indels 282;	Gaps 44;

[illegible]

```
Db 474 AORSM-----FOTIKOLEORTILQANIR-WQOEELHIQEOCLVODSNVOMFLQOP 527
Oy 621 V-----SSERADGSRHDSKGOTKLLQLLTYSKSDQMEPPSLASLSDTNKSDTSGSLPGSG 676
Db 528 VLSFSSSIORPAQOQLOQRAAQPOLVO-----LOGOI 559
Oy 677 STGTSTLKEKHKLHLRLDSSSPVDLAKLTATGKDLQSOESSSTAPGSEVITIKQEPVS 736
Db 560 ST-----QVTOHLRRESSV---ISOGPKPMRSSOLSGRSSLSLSPFSST-----599
Oy 737 PKKENALLRYLLDKDPTKIGLPEI--TPKLERLSDKTDPAS--NTKLIAMTKEREKMS 792
Db 600 -----LPLLTTPASTPPODSQCPSPDFHQRRLLSQPIOP 637
Oy 793 FEFGD----QPGSELNLEIIDLQNSQLPOLPPTRGAPAGSVKQAIINDLMQLTA 848
Db 638 MMGSGDARQP--SEVSTGQVRYAQ--SQPPD--HPNSSPVLIMG---QAVLHP--SFPA 688
Oy 849 ENSPVTPVGAOK-----TALRISQ-----STFNNRPGQGRLLPNQNLPLDITL-QS 895
Db 689 SPSPLOPAQOQOPPOAPTSLHSEDSLSLSTFSQ-QPGLGYOOP-QPRPRVSLSES 746
Oy 896 P 896
Db 747 P 747

RESULT 26
US-08-885-291-51
; Sequence 51, Application US/08885291A
; Patent No. 6051125
; GENERAL INFORMATION:
; APPLICANT: Takahashi, Joseph S.
; APPLICANT: Turek, Fred W.
; APPLICANT: Pinto, Lawrence H.
; TITLE OF INVENTION: CLOCK GENE AND GENE PRODUCT
; FILE REFERENCE: 0290-5
; CURRENT APPLICATION NUMBER: US/08/885, 291A
; CURRENT FILING DATE: 1997-06-30
; EARLIER APPLICATION NUMBER: 08/816, 693
; EARLIER FILING DATE: 1997-03-13
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 51
; LENGTH: 747
; TYPE: PRT
; ORGANISM: Mus musculus
US-08-885-291-51

Query Match 3.5%; Score 267; DB 3; Length 747;
Best Local Similarity 20.9%; Pred. No. 6,7e-11;
Matches 201; Conservative 133; Mismatches 345; Indels 282; Gaps 44;
```

```
Oy 4 MGENTSPSAETFRKKKECPDQLGSPSKRNTKRNROENKTYELAEILFANFNIDNF 63
Db 1 MDDEKXRARASRKNK-----SEKKRRDQFVVLKELSSMLPGNIR-----41
Oy 64 NRPDKCALIKETVKOIRKBOEKAAANIDEVOKSDVSTGGGVIDKALGPMLEAL 123
Db 42 --KMDKTTVEKYGIFLQ--KHNEVSAQETICDQODWKS-----FISNEEFTQMLEAL 93
Oy 124 DGEFFVNLGNVVESENTOYLRVNOEELMNSVSYLHVGDTEFVNLPLKSTIVG 183
Db 94 DG-FIAVTGTGSIYVSDSTPLGLHLPSPDVMOQNLNLPDEGHSHVYLLSHMLVTD 152
Oy 184 GSNVSGEPPRRNSHTFNCRLMKPLPDSEEGHD-----NOEAKQKTYMQC---FAVS 233
Db 153 SPSPVYLKSNDDLEFYCHLRLGSLNPKPEPTYEYIKFVGNFRSNVNPSPSCMGFNDTLS 212
Oy 234 QPSKIRGEGLDQCLICVARVP---MKERPVLPSESEFTTRQDLOGKITSIDTSTMA 290
Db 213 RPCRIVPLG---KVCFTATVYLAATPOFLKEMCYDEPLEEFTSRHSLEKFLFD---HRA 265
```

```
Oy 291 AMKPGM---EDLVRCIOKFAHOGEESVYAKRHHEVLRQGLARSQIYRSLSDGTV 347
Db 266 PPIIGLPEFVEVLETSGYDYVHI-----DLELLARQHONLMQFGKSCCYRFLTKQOWI 321
Oy 348 AAQTKSKLRSQTNRPQVYLSH-----MLHREONVCVNNPDLTGQTMKPLMPTS 399
Db 322 WLQTHYIITYHOMNSKPEFLVCTHVSVAADVAVERRQELALBDP-----366
Oy 400 SNSPAHQALCSGNPGODMTLSSNINFPINGPKRQOMPMGRFGSGSGMNVHVSQMAATPO 459
Db 367 --PEAISA-----KDDSLSE-----PRQ-----FALDQASGLSP- 394
Oy 460 GSNVYALKMNSPSCSSPGMNPQOPTSLSPRHNSPCVAGSPRI---PSQSPAGSLHS 515
Db 395 -----SPSASSRSSHKSHTAMSEPISTPTKIMASTALPRATLPOEL-----437
Oy 516 PVGVCSSSTGNSHSYTSSLSNALQALSEGHCYSLGSSLASLPDLKMGLOLNSPVMMNPPLS 575
Db 438 PVG-----LSQAATMPPLSSSC---DLTQQLQPOLQSPAPQPS 473
Oy 576 KMGSLDKOCFGLYGPSEGTGQAESSCHPGEOKE-----TNDPNTL-----PPA 620
Db 474 AORSM-----FOTIKOLEORTILQANIR-WQOEELHIQEOCLVODSNVOMFLQOP 527
Oy 621 V-----SSERADGSRHDSKGOTKLLQLLTYSKSDQMEPPSLASLSDTNKSDTSGSLPGSG 676
Db 528 VLSFSSSIORPAQOQLOQRAAQPOLVO-----LOGOI 559
Oy 677 STGTSTLKEKHKLHLRLDSSSPVDLAKLTATGKDLQSOESSSTAPGSEVITIKQEPVS 736
Db 560 ST-----QVTOHLRRESSV---ISOGPKPMRSSOLSGRSSLSLSPFSST-----599
Oy 737 PKKENALLRYLLDKDPTKIGLPEI--TPKLERLSDKTDPAS--NTKLIAMTKEREKMS 792
Db 600 -----LPLLTTPASTPPODSQCPSPDFHQRRLLSQPIOP 637
Oy 793 FEFGD----QPGSELNLEIIDLQNSQLPOLPPTRGAPAGSVKQAIINDLMQLTA 848
Db 638 MMGSGDARQP--SEVSTGQVRYAQ--SQPPD--HPNSSPVLIMG---QAVLHP--SFPA 688
Oy 849 ENSPVTPVGAOK-----TALRISQ-----STFNNRPGQGRLLPNQNLPLDITL-QS 895
Db 689 SPSPLOPAQOQOPPOAPTSLHSEDSLSLSTFSQ-QPGLGYOOP-QPRPRVSLSES 746
Oy 896 P 896
Db 747 P 747
```

```
RESULT 27
US-09-496-672-51
; Sequence 51, Application US/09496672
; Patent No. 6291429
; GENERAL INFORMATION:
; APPLICANT: Takahashi, Joseph S.
; APPLICANT: Turek, Fred W.
; APPLICANT: Pinto, Lawrence H.
; TITLE OF INVENTION: CLOCK GENE AND GENE PRODUCT
; FILE REFERENCE: 0290-5
; CURRENT APPLICATION NUMBER: US/09/496, 672
; CURRENT FILING DATE: 2000-02-03
; PRIOR APPLICATION NUMBER: 08/885, 291
; PRIOR FILING DATE: 1997-06-30
; PRIOR APPLICATION NUMBER: 08/816, 693
; PRIOR FILING DATE: 1997-03-13
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 51
; LENGTH: 747
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-496-672-51
```

Query Match 3.5%: Score 267; DB 4; Length 747;

Best Local Similarity 20.9%: Pred. No. 6,7e-11;

Matches 201; Conservative 133; Mismatches 345; Indels 287; Gaps 44;

```

OY 4 MGEINTSDPSRAETRRKRECPDOLGSPKRNTEKRNREONKYIEELAEILFANFNDINF 63
DB 1 MEDEKDRAKRASRNK-----SEKKRRDQNVILKELSMPLPGIR----- 41
OY 64 NEKPKCALIKETVKOIROIKOEKAAANIDEVOKSDVSTGGQVIDKDALGPMLEAL 123
DB 42 --KMDITVLEKYGFLQ--KHNEVSAQTEICDIOQDMKPS---FLSNEEFTQMLEAL 93
OY 124 DGEFFVNLGNVNFVSENVTOYLRYNOEELMKSVSYSLHNGDHEFFKNLPKSIYNG 183
DB 94 DG-FIAVTTDGSITIVYSDITPLGLHPSDVMQNLNLFPEOEHSEVYKILSSHMLVTD 152
OY 184 GSWSGEPFRNRNHTFNCRMLVPLPDSEEGHD-----NOEAHQYETMQC---FAVS 233
DB 153 SPSPEYKSDNDLEFYCHLRGSLNKEPTEYIKFYVGNFRSYNNVPSPSCGFONTLS 212
OY 234 QPKSIKEGEDLOSLICVARRVP---MKERVLPSSSEFTTRDIOGKITSLDTSTMA 290
DB 213 RRCRVPILG---KVCFIATVRLATPOFLKEMCVDEPLEEFTSRHSLEMKFLFD--HRA 265
OY 291 AMKPGM---EDLVRCIOKEFHNOHGESESVYAKRRHHHEVLRGGLAFSQTIRPSLSDCTIV 347
DB 266 PRTITILPREVLGTSGYDYHT---DLELLARQHILMORFGKSCCYRFLTKGQOMI 321
OY 348 MAOTKSLRSQTNPEPOLVISLH-----MLHREQNVCMNPDLQTGMKPLNPIS 399
DB 322 WLOTHYITTHQMNKSKPEITVCTHSVYADYRVERROLALEDP----- 366
OY 400 SNSPRHQALCSGNPCQDMTLSSNINPPLNGPKQKMGKMGKRGSGGMHNVSGMQATTPQ 459
DB 367 --PEHNSA-----KKDSLE-----PRO-----FALGASGLSP- 394
OY 460 GSNVALKMNPSPOSSPGMNPQOPTSMLSPRHMSPGVAGSPRIP---PSOPSPAGSLMS 515
DB 395 -----SPASSRSRSHKSSHTAMSEPISTPTKILMASTALDPRATLPOEL----- 437
OY 516 PVGCSSTGNSSHSTYNNLSINALQALSEGHVSLGSSLASPLDKMGNLQNSPYVMNPPLS 575
DB 438 PVG-----LSQATMPLSSSC---DLTQQLQOTQLOSPAPQPS 473
OY 576 KKGSLDKDCFLGYEPSEGTGGAESSCHPEQKE-----TNDPNL-----PPA 620
DB 474 AOFSM-----FOTIKDQLEORTIRILQANIR--WQOELHKTIOEQCLVODSNVOMFLQOPA 527
OY 621 V-----SSERADGOSRLHOSKQOTKLLDILTTSQDMERPPLASLSLDTKKDSTGSLPGSG 676
DB 528 VLSFSSSTIORPAQOOLQORAAQPOLVQ-----LQGGOI 559
OY 677 STHTGSLKEKHILHRLLODSSSPVDLAKLTAETGKDLQSOESSSTAPGSEVTIKQEPVS 736
DB 560 ST-----QVTHLHRESSV---ISGPRKMSQSLSGSSSLSPSSST----- 599
OY 737 PKKKENALLRYLLDKDVTKIDGLPEI--TPKLERLDSKTDPAS--NTKLIAMKTEKEENS 792
DB 600 -----LPPLLTTPASTPDQSQCPSPSFHDRLRLLSLQPIOP 637
OY 793 FERPD-----QPSSELDNLEIILDDQNSQLPOLFPDTRGAGASVDKAIINDMLQTLA 848
DB 638 MMGSGCDARQP--SEVSRTGROYKYAO--SOFPD--HPNSSVLLMG---QAVLHP--SFPA 688
OY 849 ENSPTPVGAOK-----TALRISQ-----STFNNRPGOLGRLLPNOLPLDITL--OS 895
DB 689 SPSPIQPMQAQOQPPRQAPATSLHSEQDLSLSTFSQ--QGTGLGYOQP--QPRRRRVSLSES 746
OY 896 P 896
DB 747 P 747

```

RESULT 28

US-08-885-291-52

; Sequence 52, Application US/08885291A

; Patent No. 6057125

; GENERAL INFORMATION:

; APPLICANT: Takahashi, Joseph S.

; APPLICANT: Turek, Fred W.

; APPLICANT: Pinto, Lawrence H.

; TITLE OF INVENTION: CLOCK GENE AND GENE PRODUCT

; FILE REFERENCE: 0290-5

; CURRENT APPLICATION NUMBER: US/08/885,291A

; CURRENT FILING DATE: 1997-03-30

; EARLIER APPLICATION NUMBER: 08/816,693

; NUMBER OF SEQ ID NOS: 55

; SOFTWARE: Patentln Ver. 2.0

; SEQ ID NO 52

; LENGTH: 824

; TYPE: PRT

; ORGANISM: Mus musculus

; US-08-885-291-52

Query Match 3.5%: Score 265; DB 3; Length 824;

Best Local Similarity 20.7%: Pred. No. 1.1e-10;

Matches 203; Conservative 143; Mismatches 391; Indels 244; Gaps 43;

```

OY 4 MGEINTSDPSRAETRRKRECPDOLGSPKRNTEKRNREONKYIEELAEILFANFNDINF 63
DB 1 MEDEKDRAKRASRNK-----SEKKRRDQNVILKELSMPLPGIR----- 41
OY 64 NEKPKCALIKETVKOIROIKOEKAAANIDEVOKSDVSTGGQVIDKDALGPMLEAL 123
DB 42 --KMDITVLEKYGFLQ--KHNEVSAQTEICDIOQDMKPS---FLSNEEFTQMLEAL 93
OY 124 DGEFFVNLGNVNFVSENVTOYLRYNOEELMKSVSYSLHNGDHEFFKNLPKSIYNG 183
DB 94 DGFIATVTTDGSITIVYSDITPLGLHPSDVMQNLNLFPEOEHSEVYKILSSHMLVTD 153
OY 184 GSWSGEPFRNRNHT---FNCRMLVPLPDSEEGHD-----NOEAHQYETMQC---F 230
DB 154 ---SPSEYKLSQSDLEFYCHLRGSLNKEPTEYIKFYVGNFRSYNNVPSPSCGFONTLS 210
OY 231 AVSQKSIKEGEDLOSLICVARRVP---MKERVLPSS--ESFTTRDIOGKITSLDTS 286
DB 211 TLSRPCRV--PLKEV--CFIATVRLATPOFLKEMCIYDEPLEEFTSRHSLEMKFLFD-- 265
OY 287 TMRAMKPGM---EDLVRCIOKEFHNOHGESESVYAKRRHHHEVLRGGLAFSQTIRPSLSD 343
DB 266 --HRAPTIIGYLRPREVLGTSGYDYHT---DLELLARQHILMORFGKSCCYRFLTKG 320
OY 344 GTLVAAOTKSKLRSQTNPEPOLVISLH-----MLHREQNVCMNPDLTG-----O 389
DB 321 QQMIMVQTHYITTHQMNKSKPEITVCTHSVYADYRVERROLALEDPSEALHSAK 380
OY 390 TWGKPLNP-----ISSNSPAHQALCSGNPCQDMTLSSNINPPLNGPKQDMG 435
DB 381 DKSSLEPRQHNALDVGASGLNTHSPASSR--SSKSSHTAMSEPISTPTKILMAEAST 439
OY 436 MEMGRFGSGGAMNHVSGMATTPQ-----GSNVALKMNPSPOSSPGMNPQOPTSMLSPRH 490
DB 440 PALPR-----SATLPOELPVGLSOAATMPAPLPS-----PLSCDLTQ 478
OY 491 RMSPGAASPRIPPSOFSAGSLSPVGVCSSTGNSSHSTNSLNALOA-----LSEGG 545
DB 479 LRPQVYLGSTPRAPMAQFSQFSHPOTI-----KDLBQRTIRLQANITMOEELHK 529
OY 546 VSLGSSLASPLDKMGNLQNSPYNM--NPPLSKMGSLDKDCFLGYEPSEGTGGAESS 603
DB 530 IOEQLCL-----VDSNVOMFLQOPAVSL-----SFSSIORPQAQOOLQORSA 572

```

QY 604 CHPEQKETNDPMLPPAVSSERADGSRHDSKQTKLQLLTTKSDOMEPSPLASS-LS 662  
 Db 573 AVTQPOLGAG-POLPGQISSAQVTSQHILRES-----SVISIQG----PKPMRSSQLM 620  
 QY 663 DTNKDSTGSL--PSSGSTHGTSLKEKKHILHRLQDSSSPVDLAKLTAETGKDLSESS 720  
 Db 621 QSSGRSSSSLSVSPSSAT-----AALPSSLTLTTPASTSODASQ--- 659  
 QY 721 STAGSEVTIKOEPVSPKKENALLRYLDDKDTKIDGLPEITPKLERLSDTPA--SN 778  
 Db 660 -----CQP-SPDFSHDRQLRLLSQ-----PIQPMGSCDARPESEVR 698  
 QY 779 TKLIAMTEKEEMSEFEGDQPGSELNLEITLDDQNSQLPQ-----PDDTRPG 828  
 Db 699 TGRQVKYAAQSQTVEQNDAPHANS-----SSAPMVLMLGQAVLHSPFASQPS 747  
 QY 829 APAGSVKQALINDMLQTLAENSVPYTVGAOKTALRISQSEFNPRRGOLGRLLPNOLP 888  
 Db 748 PLOPAQARQOPPOHYLOVA---PTSLHSEQDQSLLL--STYSO-QPCTLGYPQPPAP 801  
 QY 889 LDITLQSPGTGAGPPPIRNS 909  
 Db 802 QPLR-----PPRRVSS 812

RESULT 29  
 US-09-496-672-52  
 ; Sequence 52, Application US/09496672  
 ; Patent No. 6291429  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Takahashi, Joseph S.  
 ; APPLICANT: Turek, Fred W.  
 ; APPLICANT: Pinto, Lawrence H.  
 ; TITLE OF INVENTION: CLOCK GENE AND GENE PRODUCT  
 ; FILE REFERENCE: 0290-5  
 ; CURRENT APPLICATION NUMBER: US/09/496,672  
 ; CURRENT FILING DATE: 2000-02-03  
 ; PRIOR APPLICATION NUMBER: 08/885,291  
 ; PRIOR FILING DATE: 1997-06-30  
 ; PRIOR APPLICATION NUMBER: 08/816,693  
 ; PRIOR FILING DATE: 1997-03-13  
 ; NUMBER OF SEQ. ID NOS: 55  
 ; SOFTWARE: Patentln Ver. 2.0  
 ; SEQ. ID NO 52  
 ; LENGTH: 824  
 ; TYPE: PRT  
 ; ORGANISM: Mus musculus  
 ; US-09-496-672-52

Query Match 3.5%; Score 265; DB 4; Length 824;  
 Best Local Similarity 20.7%; Pred. No. 1.1e-10;  
 Matches 203; Conservative 143; Mismatches 391; Indels 244; Gaps 43;  
 QY 4 MGEHTSPSRAETRRKKECPDOLSPSKRNTKRNREOKNTYIELALIFANNDIDNF 63  
 Db 1 MDEDEKARAKASRNK-----SEKKRQDFNLKELSMLEGNR----- 41  
 QY 64 NFKDKCAILKETVKQRIKEEKAANAANDEYOKSDVSSTGGVIGDKDLGPMLEAL 123  
 Db 42 --KNDKTTVEKVIQGLQ--KHNEVSAQTEICDIOODKPS-----FLSNEFTQMLLEAL 93  
 QY 124 DGEFFVNLQGNVVESENVTQYLRNQDELANKSVYSLHVGCHTEFVNKLPRKSTVNG 183  
 Db 94 DGFIAVTTGSGSIYVDSITPILGHLPSDVMDQNLNLFPEQHSVEYKILSSHMLVTD 153  
 QY 184 GSNMGSEPPRRNSRT--FNCMLVYKPLPDSSEEGHD-----NDAHOKYETMQC----F 230  
 Db 154 ---SPSEYILKSDSDLEFYCHILGSLNPKKEFPYIEYIKFVGNRNNVPPSCNGEDN 210  
 QY 231 AVSQPKSITKEGEDLQSLICIVARRVP--MKERPVLS--ESFTTQDLOGKTTISLDT 286  
 Db 211 TLSRPRCV-PLGKEV--CFIATVRLATPQFLKMCIVDEPLEEFTSRHSLEMKFLFD-- 265

QY 287 TMRAMKPGW---EDLYRRCIOKFAHGESEVYAKRHHNEVLRQGLAFSQIYRESLSD 343  
 Db 266 -HRAPPIIGVLPFEVLGTSGYDYHI---DDELLARCHQHLMQFGIGSCCYRFLTKG 320  
 QY 344 GTLYAAQTCKKLRSQTTNEPOLYISLH-----MLHEQYVCVNNPDLTG-----Q 389  
 Db 321 QQWIMLOTHTYIYHWNKSPREFVCTHVSADVAVVERROELALDEPSEALHSALK 380  
 QY 390 TMCKPLNP-----ISSNPAHQALCSGNPGQDMTSSNINPINOPEQMG 435  
 Db 381 DKGSLEPRQHFNALDVGAGLNTSHPSASSR--SHKSSHTAMSEPTSTPTKLMAEST 439  
 QY 436 MPMGFRGSGGMNHVSGMQATTPQ-----GSNVALKNSSPQSGNPGQPTSMLSPRH 490  
 Db 440 PALPR-----SATLPQELFVPGISQAATPAPLPS-----PLSCDLTQ 478  
 QY 491 RMSGVAGSPRIPRSQSPAGSLHSPGVCSSTGNSHTYNNSSLMALQ-----LSRGH 545  
 Db 479 LLPQTVLQSTPAPMAQPSAQSMTORT-----KDLEORTRLIQAIRMQOEELH 529  
 QY 546 VSLGSSLASPDLKMGKQNSPVNM--NPPLSKMSLSDKDCFLYGESEGTGQAES 603  
 Db 530 IQEQLCL-----VQDSNVQMFLOQPAVSL-----SPSSIORPRAQOQLQGRSA 572  
 QY 604 CHPEQKETNDPMLPPAVSSERADGSRHDSKQTKLQLLTTKSDOMEPSPLASS-LS 662  
 Db 573 AVTQPOLGAG-POLPGQISSAQVTSQHILRES-----SVISIQG----PKPMRSSQLM 620  
 QY 663 DTNKDSTGSL--PSSGSTHGTSLKEKKHILHRLQDSSSPVDLAKLTAETGKDLSESS 720  
 Db 621 QSSGRSSSSLSVSPSSAT-----AALPSSLTLTTPASTSODASQ--- 659  
 QY 721 STAGSEVTIKOEPVSPKKENALLRYLDDKDTKIDGLPEITPKLERLSDTPA--SN 778  
 Db 660 -----CQP-SPDFSHDRQLRLLSQ-----PIQPMGSCDARPESEVR 698  
 QY 779 TKLIAMTEKEEMSEFEGDQPGSELNLEITLDDQNSQLPQ-----PDDTRPG 828  
 Db 699 TGRQVKYAAQSQTVEQNDAPHANS-----SSAPMVLMLGQAVLHSPFASQPS 747  
 QY 829 APAGSVKQALINDMLQTLAENSVPYTVGAOKTALRISQSEFNPRRGOLGRLLPNOLP 888  
 Db 748 PLOPAQARQOPPOHYLOVA---PTSLHSEQDQSLLL--STYSO-QPCTLGYPQPPAP 801  
 QY 889 LDITLQSPGTGAGPPPIRNS 909  
 Db 802 QPLR-----PPRRVSS 812

RESULT 30  
 US-08-816-693A-52  
 ; Sequence 52, Application US/08816693A  
 ; Patent No. 5874241  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Takahashi, Joseph S.  
 ; APPLICANT: Turek, Fred W.  
 ; APPLICANT: Pinto, Lawrence H.  
 ; TITLE OF INVENTION: CLOCK Gene and Gene Product  
 ; NUMBER OF SEQUENCES: 53  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Dressler, Rocky, Milnamow & Katz  
 ; STREET: Two Prudential Plaza, Suite 4700  
 ; CITY: Chicago  
 ; STATE: Illinois  
 ; COUNTRY: USA  
 ; ZIP: 60601  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patentln Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:



```

OY 124 DGEFVYVNEGNNVENVSEVTOYLRYNOBELANKSVYSIIHVGDTHEFYKNLLPKSIVNG 183
DB 94 DGEFIIVTTDGSIIYVSDSTPLILGLPSPVMDQNLINLPEOEHSSEVYKILSSHMLVTD 153
OY 184 GMSGSGPPRRNSHT---FNCRMLVKPLPDSEEGHD-----NOEAKHOKETIMOC-----F 230
DB 154 ---SPSEPIKSSDLEFCHLRLSGSLNKEPFTYEIKFVGNFRSTNNVSPSCGFDN 210
OY 231 AVSQPSIKEEGEDLOSCLICVARRVP---MKERPVLPSS-ESFTTODLOGKITSLDTS 286
DB 211 TLSRCPRV-PLGKEV--CFIATVRLATPOLKEMCIYDELEEFTRSHLSLEMKFLFLD-- 265
OY 287 TMRAMKPGM---EDLVRCIQKFHQAHEGESVYAKRHHEVLRQGLAFSQTIRPSLSD 343
DB 266 -HRAPPIIGLPEPEVLGTSGYDYH---DDELLARCHOHLMOGTGKSCCYRFLTKG 320
OY 344 GTLVAQTSKILRSQTTNTPOLVLSHMLHREQNVCMNPDLTGQTMKFLNPISNSP 403
DB 321 QOMIWLQTHYITYYHOMNSKPEFTVCTHVSYSTADVRV-----E 339
OY 404 AHOALCSGNPGQDMLTSSNINFPINGKEQMGPMGRFG---GSGGMNHVSGMOATTPQ 459
DB 360 RQGLALDEPPSEALHSSALK-----DKGSSLPRQHFNALDVGASGLN----- 403
OY 460 GSNYALKMNSPSSQSPGMANGOPTSMLSPRHRMSP--GVAQSPRIPPSQSPKSLHSPY 517
DB 404 ---TSHSPSASSRSSHKSSTHMTAMSEPTSTPTKLAESTPALPRSATLPQ---ELPV 454
OY 518 GVCSTGNSHTYSSNLNALQALSEGCVSLGSSLASP---DLKMGNLQSPVNMNPPPL 574
DB 455 -----PGLSQ--AATWPAPLPSPLSCDLTQOLLPOTVLQSTPAPM 492
OY 575 SKMGSLSKDCFGIYGEPESEGTGGAESSCHPEQOKETDNPPLPVVSSERAD---GOSR 631
DB 493 AGPSAQ---FEMF-----QTIKQLEQRTIRILQANIRMOQEE 526
OY 632 LHSKQOTKILQLLTKSDOMERSPLASSLSDTNKSDTSGSLPGSGTHGTSLEKHKILH 651
DB 527 LHKIOBOLCLVODSNVOMFLQPA-VLSFSSSTORPEA-----O 564
OY 692 RILQDSSSPVDLAKLTAEATGKDLQSSSTAGSEVYIKOEPV---SPKKENALLRY 747
DB 565 QOLQOQSAVTPQLGA---GPOLPGOISSAQVTSOHLHRESSYISTOGKPKMRSSQLQ 621
OY 748 -----LLDKDDTKDIGPEITPKLERLSDXTDPASNTKLIAKTEKEMSFEEDOP 799
DB 622 SSGRSSSIVSPSSATAALP---PSL---NLTPASTS-----OD 656
OY 800 GSELNDLEETLDLQNSQLPQLEFPDTRPGAPASVDKQAIINDLMQUTAEVTPVGAQ 859
DB 657 ASQCOPSPDSHRODLRL--LSQPIOPMP--GSCDAR-----OPSEVSRTGRQ 702
OY 860 KTLARISOSTFNNPRPGQLRLPNQNLPLDITLQSTGAGPPPIRNSPYVIPOGM 919
DB 703 VKTAQ--SQYFQNP-----DAHPANSSAPMYL----- 730
OY 920 MGNQMGINGNGLNSGTGMIASASRPTMPSGEMAPQSSAVRYTCAATTSAANRPYQG 979
DB 731 -----LMGQAVLHPSPASQ----- 745
OY 980 MIRNPASTIPMRPSQGOQOTLOSQVYMINGPSELEKMMGGOYSQO-----QAPPNQ 1032
DB 746 ---PPLPQPAQARQOPQHY-LQVQAPYSLHSEODDLSLSTYSQOPGTIGYPOPPAQ 800
OY 1033 TAPWPSIILPIDQAFASON---ROP 1055
DB 801 ----PQPLRPPRRVSSLSESSGLQOP 822

```

RESULT 32  
 US-08-061-376-5  
 ; Sequence 5, Application US/08061376  
 ; Patent No. 6175000

```

; GENERAL INFORMATION:
; APPLICANT: Evans, Glen A.
; APPLICANT: Djabbali, Malek
; APPLICANT: Selleri, Lucia
; APPLICANT: Parry, Pauline
; TITLE OF INVENTION: CHARACTERIZATION OF A CHROMOSOME 11023
; TITLE OF INVENTION: TRANSLOCATION BREAKPOINT ASSOCIATED WITH ACUTE LEUKEMIAS
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESSES:
; ADDRESS: Pretty, Schroeder, Brueggemann & Clark
; STREET: 444 South Flower Street, Suite 2000
; CITY: Los Angeles
; STATE: California
; COUNTRY: USA
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/061,376
; FILING DATE: 13-May-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Reiter, Stephen E.
; REGISTRATION NUMBER: 31,192
; REFERENCE/DOCKET NUMBER: P41 9387
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619)546-4737
; TELEFAX: (619)546-9392
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3969 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-061-376-5

Query Match 3 4%; Score 258.5; DB 4; Length 3969;
Best Local Similarity 18.6%; Pred.No.3.9e-09;
Matches 350; Conservative 243; Mismatches 611; Indels 673; Gaps 91;

OY 10 DPSRAETRRKKECPDOLGSPKRNTERKRNREQENKYIEELAEILFANFNIDINDNFNPKPK 69
DB 1672 NPETESIPRSRSP--GPPPVLTGVSKDDQDPL-----DLGVKRRKMQ 1716
OY 70 ---CALIK--ETVKQIRQIKEDOKAAANID---EVOKSDVSSTGQGYIDKDALGPM 119
DB 1717 GNYTSVLEFSDDIVKILQ-----AATNSDCGOPEIKKAN-----SM 1752
OY 120 LEALDGEFFVYVNEGNNVENVSEVTOYLRYNOBELANKSVYSIIHVGDTHEFYKN--LLPK 178
DB 1753 VKS---FLLRQMERVFPWFSVAKSRWEPNK-----VSSNSGMLPNNAVLP 1795
OY 179 SI-VNGSGWSGPPRRNSHTFNCRLVYKPLPDSEEGHDNOEAHOKETIMOCFAVSQPS 237
DB 1796 SLDHNTAQW--QRENSHTQEPPLAKKTIIPAKRPGRGPPDSPTPLHPTPTLTSDS 1853
OY 238 IKEEG-----EDLOSCLICV-----ARRVPKERP-----VLPSSSEFTTR 273
DB 1854 REDSPELNPPGIEDNRQCALCLTYDDDSANDAGRLLYIGQNEWTIVNICALMSAEVF--- 1910
OY 274 QDLOGKITSLDSTMA-----AMKPGWE--DLVRCIQKFHAQHE-----GESV 316
DB 1911 EDDGSLKNVHMAVIRGKQLRCEFCOKPGATVGCCLTCSNTHFMCSTRAKNCVFLDKK 1970
OY 317 SYAKRHH-----HEVLRQ-----GLAFSQIYRF----- 339
DB 1971 VYQHRHDLIKGEVPENGEVEFRVVDPEGISLRKRKLNGLEPNIHMMIGSMITDCL 2030

```

```

Qy 340 ----SLSDGTLVAAQTK-----SKLIRSQTNEPOLVILSHLHREOVNCPANPLDT 387
Db 2031 GILNDLSD-----CEDKLEPIGYCSRYWMTTDARKRCVYTCCKIV--ECRPVVEPDI- 2082
Qy 368 GQTMCKPLNPISSPAHQALCSGNPGODMLSSNINFPINGPKQMGKMPKRGCSGGM 447
Db 2083 NSTVHEDNRTIAHSPFTFTESSSKESQNTA-----EIIISPSPDRPHSQTSGSCYV 2135
Qy 448 NHVSGM-OATPGSGNVALKNNSPSSQSGMNP-----GQPSMLSRRH-----MSP 494
Db 2136 HVISKVPKIRTPS-----YSPQRSPGCRPLPSAGSP--PTTHEIVTVGDPILSS 2184
Qy 495 GV--AGSPRIPSPSPSPAGS---LHSPVGVCSSTGNSHTYSS-----533
Db 2185 GLRSTGSRHRSSTLSLSPKSKRIMSPM---RIGNTYSRRNVSSVSTGTATDLESSAK 2240
Qy 534 -----LMLQALSEHGVSLSGLASPLDKMGNLONSPVNNMPPILSKWS---LDS 582
Db 2241 VVDHYLGLPNSSTSL--GQNTSTSSNLQRTVTVYVGN-KNSHLDGSSSEMKOSSASDIYS 2297
Qy 583 KDCFLGYE-----PSEGTQOAESSCHPGQKEINDNLPRAVSEERADGSRLLHDSKG 637
Db 2298 KSS-SLAGEKTKVLSKSSBESAHNAVPG-----IPKLAP-----2332
Qy 638 QTKLQLLTTRKSDQMEPSPLA-----SLSDTNKSDSTGLSGSGTGTSLKEKHILHR 692
Db 2333 -----QVHNTTSRELNVSKISFAPSSVSFSKE-----ALSFLHLRG 2373
Qy 693 LLQSSSPVDLAKLTAEYTGDLDSOESSYAPGSEVYTKQEPVSPKKENALLRYLLKD 752
Db 2374 QNRNRDQHTD-----STOSANSSPDEDEYEVTKLKSQMSNRSSII-----2413
Qy 753 DTKDGLPEITPKLERLDSKTPASNTKLIAMKTEKEEMS---PEPDQCGSELDNLE- 807
Db 2414 -NEMHG-----SSSDRROKGGKSKCKEYFEKHSSKSLPEQVTTGEGGNLKP 2461
Qy 808 EILDLQNSQLPOLFPD-----TRPGAP-----AGSYVD-----835
Db 2462 EFMDEV-----LTPEVMGQRPCNNVSSDKIGDKLSMPGPKAPPMQVESAELQAP 2514
Qy 836 -KQALINDLMQJLAEV-----SPVTPVGAQKATLRISOSTFNNPRPQGLRLLP 883
Db 2515 RKRRTVYKTLTFLKMEENESQKNALEKSSPASPLQIESTSPTEPISASENPGDGYAQPSP 2574
Qy 884 N-----ONLPL-DITLQSPTAGP-----PP-----PIR 906
Db 2575 NNTSCQDSQSNNTQNLPLPQDNLMLPDGPKPOEDGSEFKRRIRPRRSARARSNNFGLTPLY 2634
Qy 907 NSSPY-----SVIPOPGMGNQGMIGNQNLGNS-----STGMIG 941
Db 2635 GVRXYGEEDIPFYSSSTGKRGKRSABEQVVGADDLSTSDDDLXYNYFTRTVYSSGEE 2694
Qy 942 NSASR-----PTMPSGEMAPSSAVRVTCATTSAMNR-PVOGGMIRNPAASIP 989
Db 2695 RLAHNLFREREQCDLPKISQLDVDDGTSTSVTATFRKSSQIPKRNKG-ENGTENILK 2753
Qy 990 MRPSQSGQROTLQ-----SOVMNITGSELENNMGPOYSQ-----1025
Db 2754 IDREPDAGEKEHVTKSSVGHKNEPKMNCCHSVRYKTOGGQDLSQLSLESRRVHTST 2813
Qy 1026 -----OQAPNPQTAWPESILPIDQASFASON---ROFGSSPDLLCLPH 1067
Db 2814 PSDKNLDTYNTLLKSSDSDNNNSDGCNILLPSDIMDFVLKNTPMQALGESP-----2866
Qy 1068 PAASPSDEGALLDQLYALAEFNQGLE-EIDRALGIPELVVSQSAVDEQFSSQDSNIML 1126
Db 2867 -----ESSSSE-----LNLLEEGGLGIDSNREKDMGLEFVFSQQLPTTEPVSSVSSISA 2916
Qy 1127 EQ-----KAPVFPQOYASQAQMAQGSYSMPMODNFTTMQR-----P 1163
Db 2917 EEOGELPELEPSDLSVLTTRSPYVPSQNPSPRLAYISDS-----GKKRVYTIK 2964
Qy 1164 SYATLRMPQ---RGLRPTGLVQNOPQLRLQLOHRLQAOQONROPILMNOI--SNVSNVNL 1218

```

```

Db 2965 SVASSESDPALLSPGVDPDPREGHMTDPDH---IQGHMDADHITSPSCSVGEGHGNQDL 3021
Qy 1219 TLRRGVP--TQAFINMOMLAQORELINO-----HLR---QOQMHOQ 1255
Db 3022 TRNSSTPGLQYVSPSPVPLQIONKYYPNSTDSPSQISNAVAQYTPPHLKPAETKLIYVN 3081
Qy 1256 QOVQOQ-----RTL-----MMRGQGLNMTP-----SMVAPSGMPATMS---NPRIOA 1294
Db 3082 QMOPQLYVLQTLPLKCVTQKIQTLTSSVSTPSVMEETNTSVLGMGGGLTLYGLNLSLPLTS 3141
Qy 1295 N-----AOQFPPEPN-----YGISOOPDPCF-----1315
Db 3142 QSLFPSASKGLLPMHQHLSFPAAQTSSFPNPISNPPSGLLIGVQPPDPQLVSSS 3201
Qy 1316 -----TGATTPQSPPLMSPRMAHTOSPMMQO--SQANPAYQAPSDI-----1353
Db 3202 QRTDLSTTVATPPSSQIKRRPISRLQTRKXKLIAPSSFTSPSNIAPSDVSNMFLINFTPSOL 3261
Qy 1354 -NGMAQGNMGNSMFSQSPHFGQANTSS-MYSNNMNIIVSMATNTGMSMNMQMTQOI 1411
Db 3262 PNHPSLIDLGSLNTSSHRTVPNITIKRSKSIWTFEPAPL---LPQSVGTAATAAGTSTI 3318
Qy 1412 SM--TSVTSVSTSGLS 1426
Db 3319 SODTSHLTSGSVSGLAS 3335

```

```

RESULT 33
US-08-194-468-2
; Sequence 2, Application US/08194468
; Patent No. 5750336
;
GENERAL INFORMATION:
APPLICANT: Montminy, Marc R.
TITLE OF INVENTION: ASSAYS FOR THE IDENTIFICATION OF
TITLE OF INVENTION: COMPOUNDS WHICH INHIBIT ACTIVATION OF CAMP AND MITOGEN
TITLE OF INVENTION: RESPONSIVE GENES
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark
STREET: 444 South Flower Street, Suite 2000
CITY: Los Angeles
STATE: California
COUNTRY: USA
ZIP: 90071
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/194,468
FILING DATE: 10-FEB-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Reiter, Stephen E.
REGISTRATION NUMBER: 31,192
REFERENCE/DOCKET NUMBER: P41 9672
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619)-546-4737
TELEFAX: (619)-546-9392
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2441 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-194-468-2

```

Query Match 3.3%; Score 252; DB 1; Length 2441;  
Best Local Similarity 20.6%; Pred. No. 5.3e-09;  
Matches 248; Conservative 142; Mismatches 442; Indels 372; Gaps 61;





```

OY 847 TAENSPTVPVGAQKTAALRISQSTFNNPRPGQLRLLPNONLPLDITLQSPGAGFPPR 906
    | : |
    422 TRHDCPV-----CLPLKNASDKRNOQIILG----- 446
OY 907 NSSPVSVIPQGMGNOCNIGNOCNIGNSSTGMICNSASRPTMPSGEMAPQSSAVRVTCA 966
    | : |
    447 --SPAS-----GIONTIGSVGAGQOQATSL--SNPNPIDP-----SSMORAYVA 486
OY 967 ATTSAMNRYVQGMIRNPAASITPMRPSQPGOROT-LOSQVNN-IGPSELENNMGSPQYS 1024
    | : |
    487 LGLPYMNP-----OTQLOPVYGOQPPAPRAHQOQRTLNALGNMMSVPAQGITTD 538
OY 1025 QOQAPNOTAPWESILPIDQAFASQNRPF---GSSPDDL--LCPHPAESPDQGA- 1078
    | : |
    539 QO--PNLIS---ESALP---TSLGATN--PLMNGSNGNIGSLSTIPTAAPPSSTGVR 588
OY 1079 -----LLDOLYLA- RNPDGLEIDRALGIPELVSOQAVDPQSSQSDSN 1123
    | : |
    589 KGMHEVYQDLRSHLVHKLVAQIFPTPRPAALKDRRM--ENLVAYAKKEGDMYESANSR 646
OY 1124 ---IMLEOKAFVFPQOQASQOQMOGSSYPMQDNFHTMGORPSYATILRMQPRGLRPT 1179
    | : |
    647 DEYHLLAEKI-----YKIQELEKKRRTRLKQGI--LGQAPLAPSAQAP-FVITPA 697
OY 1180 GLVQNPQRLQRLQRLQAOQNRQPLMNOISVNSVNLTLR----- 1221
    | : |
    698 QSVRRNPGRPLRV-NRMOVSOG---MNSFPMASLGVOQLQOAMGRAPASPMHNSVOM 752
OY 1222 ---PGV---PTQAPINAOMLAORERILNOLRORMOQOQVOQRTLLMR---GQ 1268
    | : |
    753 NSMASVPMASISPSRMPQPRPMNMGTHANNIMQAQPTQNOFLPQNOFPSSGAMSVNSVGM 812
OY 1269 GLNMPRSVAVASGM-ATMSNP---RIPQANQOFRPPPNKISQOP----- 1311
    | : |
    813 GQPAQAQVSGOQERGAALPNPLNMLAPQAS--QLPCRP---VTQSPHPTPPRPASTAAG 867
OY 1312 ---DPGFT---GATTPQSPPLMS-----PRMATHOS-PMQO---OSQA 1343
    | : |
    868 MPSLOHPTARPGMTPRQAPAPQPTSTPVSSGOTPTPTPSVSFAAOTOSTIPTQAAQAQOY 927
OY 1344 NPAYOAPSDINGMAQGNNGNSMFSQOSP-RHFGQOANTSMYSNNMNTN---VSMATNTG 1399
    | : |
    928 TPQPTP-----VQRPVATPPOSSQOQPTPVHTQPRGRTPLSOAASIDNRVPTSTVTS 981
OY 1400 GMSMNQNTGOISWTSV--TSVTSLSLSMGRQVNDPARGNLPLNQLRPGDMKQOEBD 1458
    | : |
    982 AETSSQOQGPDPVPLEMKTEVOTDDAEPPTESKGEPR---SEMMEEDLOGSSQVKEETD 1038
OY 1459 TTRK 1462
    | : |
    1039 TTEQ 1042

```

```

RESULT 35
US-08-480-473B-4
: Sequence 4, Application US/08480473B
: Patent No. 5882914
:
: GENERAL INFORMATION:
: APPLICANT: Semenza, Gregg L.
: TITLE OF INVENTION: HYPOKIA INDUCIBLE FACTOR-1 AND METHOD OF USE
: NUMBER OF SEQUENCES: 64
: CORRESPONDENCE ADDRESS:
: ADDRESS: Fish & Richardson P.C.
: STREET: 4225 Executive Square, Suite 1400
: CITY: LA JOLLA
: STATE: CA
: COUNTRY: USA
: ZIP: 92037
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS

```

```

: SOFTWARE: Patent Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/480,473B
: FILING DATE: 06-JUN-1995
: CLASSIFICATION: 514
: ATTORNEY/AGENT INFORMATION:
: NAME: Haile, Lisa A.
: REGISTRATION NUMBER: 38,347
: REFERENCE/DOCKET NUMBER: 07265/053001
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 619/678-5070
: TELEFAX: 619/678-5099
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 805 amino acids
: TYPE: amino acid
: STRANDEDNESS: not relevant
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-480-473B-4

Query Match 3.2%; Score 248; DB 2; Length 805;
Best Local Similarity 19.3%; Pred. No. 1.8e-09;
Matches 178; Conservative 144; Mismatches 322; Indels 280; Gaps 39;

OY 39 ROENKYEIELAEELFANFNQNDNFNFKDKAILKEVYKQROJKEQEKAAAANIDEVQ 98
    | : |
    9 RSKSEVEFEELAHQPLPPLNVSRL---DKASVRLTISYLR-VKRLDAGDLDIEDMD 63
OY 99 RSDVSTGGGVIDKDALGPMLEALDGFEEVYVNLGNNVFESEVNTQYLRVQEEELMKS 158
    | : |
    64 KQAMC-----FYKALDGYVMYVLTDDGDMITISDNVKNKMGTLQFELTGS 110
OY 159 VYSILVGDHTEFVKNLLEKSTIVNGSWSGEPBRNSHTFNCRLVKPLPSEEBGHDNQ 218
    | : |
    111 VDFTHPCDHEMRMLTFR---NG-----LVK-----KGKBDN 141
OY 219 EAHQYETMOCFVASQPSIKREG-----EDLOSCLIC 251
    | : |
    142 TORSEFLRKCTVLSRGRTMNIKSAWKVLAHCTGHIHYVDINSNPOCGYKKRPMTCLVL 201
OY 252 VARRVPMKERVLP--SSEFTTRDLOGKITSLDTS--TWRAAMKRGMEDIVRCIOKPHA 309
    | : |
    202 ICEPIRHPNSNIEPIRDSKTFILSRHSLDMKFSCDERITELMKYER--BELLRSTIYEY- 258
OY 310 QHEGESVYAKRHHEVLRQGLAFSQIYRFSLSLSDGTLVAAOYTKSLINSQTNPEOLVIS 369
    | : |
    259 -HALDS-DHLTKTHHDMFTKGQVTTGOYRMLAKRGGYVVEQTAVIYNTKNSQPCIVC 316
OY 370 LH-----MLHRQNVVYMP-DITGOTMGRPLNPISNSPAH-----QAL 408
    | : |
    317 VNYVVGIIQHDLIIFSLQOTCEVLKPVESSDMKMTQLFTKVESEPTSLFDKKEPDL 376
OY 409 CSGNPGQDMTLLSNINPFLINGPKQOMGPMGRFGSGGNHVSQAQVATTPQGSNYALKM 468
    | : |
    377 TLAPAAQDTIIS-LDF-----GSNTEFDD 401
OY 469 SPSQSSPGMNGOPTSMISPRHRM--SPGYAGSPRIIPPSQFSPAGLSHPVGVCSSTGSH 527
    | : |
    402 QOLEVEVPLYN--DVMLSPNKEKONINLAMP-----LPTAETPPPL-----R 442
OY 528 SYTNSINAAOLAS--EGGVSIGSSLASPDLMGLMSVPMNMPPLSKMSLDSKDCF 586
    | : |
    443 SSADPALNOEVALKLEPPESILELSFTMOIO-----DQTP----- 478
OY 587 GLYGESEGTGQAESCHPGEOKE-----TNDNLPRAVSSERADQSRLLHDSKG 637
    | : |
    479 ---SPDGSSTNQ--SSPEPNPSSEYCYVYSDWMEKFLLEVEKLFADPEAKKPFSTQ 532
OY 638 QTKL-LQLIT-----TKSDQMEPSPLASSLSPTKNDSTGSL-----PG 674
    | : |
    533 DTDLDEMLAPYIPMDDDFOILRSFDOL--SPLESSASPEASPOSSTVTVFOQTIOIBPT 590

```







```

Db 478 MPOIDDP---ASPS-DGSTR---QSSP--EPNSPSEXC-----FDVDSMVNFK-- 519
Qy 541 SEGHSVLSGSSIASPDLKMGN-LQNSPVMMNPPLSKMSGLSDSKDQFGLYGPSEGGTGO 599
Db 520 -----LLEVERKLFADTEKKNPFSTQDPDLDEMLAPYIPMD--DDFOL----- 561
Qy 600 AESSCHPEQKFTNDPNLPPAVSSERADQSRILHDSKQFTLLQLLTTRKSDMEPSPLAS 659
Db 562 --RSFDQSLPLESNSPS--PPSMSTVTGFGQTQL-----QKPTITAT 599
Qy 660 SLSDPNKSSTGSLPSGSGHGTSLKEKHILHRLLODSSPYDLAKITPAEAT----- 711
Db 600 ATTATTTES-----KTEKDKNEKIL--IASPSTQVPOETTAKASAYSGTSHR 650
Qy 712 -----GKDSLQESSSTAPGS-----EVTIKOEVSFK--KRENALLRYLLDKXD 753
Db 651 TASPDRACKRVTEQTDKAPRLNSATLNONRYPEELNKTTIASOAKKRRKMEHDG 710
Qy 754 T--KDGLPEITPKLERLDSKTDSPASNTKLIAMKTEKEMSFEPGDQPSGLDNLEETLD 811
Db 711 SLFQAAGIGTL-----LQDPGDCAPTMSL---SMKRVKGFISSEQNGTEGKTIILIPS 760
Qy 812 DLQNSQLQLPEDTPRGAPAGSVKQAIINDLMOLTAENSPV--TVYGAOKTLR 864
Db 761 DLACRLGQ-----SMD-----VSGLPOLTSYDCEVNAPIQSGSRNLLQ 798

```

```

RESULT 40
US-08-045-806-4
: Sequence 4, Application US/08045806
: Patent No. 5378822
: GENERAL INFORMATION:
: APPLICANT: Bradfield, Christopher Alan
: APPLICANT: Dolwick, Kristin Marie
: APPLICANT: Poland, Alan
: TITLE OF INVENTION: An Receptor cDNA and Method of
: TITLE OF INVENTION: Determining Human Risks To Environmental Pollutants
: NUMBER OF SEQUENCES: 23
: CORRESPONDENCE ADDRESS:
: ADDRESS: Tilton, Fallon, Lungmus & Chestnut
: STREET: 100 South Wacker Drive, Suite 960
: CITY: Chicago
: STATE: Illinois
: COUNTRY: USA
: ZIP: 60606-4002
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/045,806
: FILING DATE: 19930408
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Feitress, Susan B.
: REGISTRATION NUMBER: 31,327
: REFERENCE/DOCKET NUMBER: NU-9207
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (312)-456-8000
: TELEFAX: (312)-456-7776
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 848 amino acids
: TYPE: AMINO ACID
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-045-806-4

```

```

Query Match 3.1%; Score 235; DB 1; Length 848;
Best Local Similarity 18.2%; Pred. No. 1.7e-08;
Matches 188; Conservative 156; Mismatches 371; Indels 320; Gaps 42;

```

```

Qy 4 MGEINTSDPSRAETRRKRECPDQAGSP-----KRNTEKRNREQENKYEIELLELI-FANFN 58
Db 1 MNSSSANITTYASRRKRKRVQKTKVPIPAEGISKNSKRRDLNLTDLALALFFP--Q 58
Qy 59 DIDNENFKPKDCAILKEVYKQIQRIKEOEKAAANIDEVOKSDVSTGCGVID-----K 112
Db 59 DVIN--KLDKLSVLRSLVSYLK-----AKSFEDVALKSSPIERNQONCRAANFR 107
Qy 113 DAL-----GPMLEALDGFPPVYVNLGCVNYFVSENTOYLTRYQOELMKSVYSLHVGDH 168
Db 108 EGLNLQEGEFLQALNGFLVLTALVFAVASTIDYLGFOQSDVYIHQSVYELLHTEDR 167
Qy 169 TEFKNLPLKPSIVNGSGNSGPEPRRNSHTFNCRLVPLRDEEGCHDNQOAKQETMQ 228
Db 168 AEFOROL-----HVALNSQ-----CTESGQIGEEATGIPQYIV 201
Qy 229 CFVVSQPSIKKEGEDL-QSLICVARRVPMKERVLPSSSEFTTRDLOGKITSL--- 283
Db 202 CY--NPQIIPENSPLMERFCIC-----RLRCLDSSGFLA-MNQGLKTYLHGQK 250
Qy 284 ---DTS-----TTRAAKPGMEDLVRCIOKFAHNE----- 312
Db 251 KKKGDGSLPLPOLALFALATPLQPPSILEIRTKNEIFFRTKHLDFTPIGCDAKGRIVLGY 310
Qy 313 -----GESVYAKRHHHEVLHQGLAFQSIQYVFSLSDTGLVAQTKSKLI 356
Db 311 TEAEICTRGSGYQFIHADMLYCASHIRMTITGESGMIVFRLTKNRMTWVQSNAR-- 369
Qy 357 RSQTTNEPOLYISLHMLHREQNVCMNPDLTGOTMGKPLNPISNSPAHQALCSGNPGOD 416
Db 370 -----LYKNGRPDYIIYQRLPDEBETENLRKNTKLP-----FMFTTGEA 411
Qy 417 MTLSSNINFP-INGPKQGMKPMKRGSGGNNHVGMOATTPOGSNTALKMNSPOSQSP 475
Db 412 VLYEATNPFPAIMDP-----LPLRTKNGTSGKD-----SATYSTLSKSL- 451
Qy 476 GMPNQPTSMLSPRHRMSPGVAGSPRIPPSQSPAGSLSPVGVCSGNSHSYNTSSLN 535
Db 452 -----NPSLLAAMQODESITLYP-----ASTYSTAPFENNFN 487
Qy 536 ALQALSEGHVSLGSSLASPDLKMGNLQNSPVMMNPPLSKMSGLSDSKDQFGLYGPSEGG 595
Db 488 --ESMNE-----CRNQDNTAPMGNDITLKHEQID-----QPD- 519
Qy 596 TTGOAESCHGEQKFTNDPNLPPAVSSERADQSRIL-----DSKQTKLL----- 642
Db 520 --VNSFAGGHPQLFQDSKNSDLYSIKMLGIDFEDIRHMQNEKFFRNDPFGVDFRIDL 577
Qy 643 --QLTTKSDQMEPSPLASSLSDTKKDSGSLPGSGSTHGTSLSKKHILHRLLODSSSP 700
Db 578 TDEILTYVQDSLKSPTIP--SDYQOQOSLAL-----NSSCMVOEHLHLEQOQHOKQ 629
Qy 701 VDLAKLTAEATGKDSLQESSSTAPGSEVTIKOEVPYSPKK-----ENALLRYL 748
Db 630 V-----VEEQOQLCQKMKHMQVNGMFENNMSNOIYFEN 663
Qy 749 LDKDDTKOI-----GLPEITPKLERLDS-----KTDRASNTKLIAMKTEKEEM-----S 792
Db 664 CPOODPOOYNWFTDLHGISOEFYKSEMDSPYTONFTSCNOVPLPOHKSKEIDLYPMGS 723
Qy 793 FEPGQPSGLDNLEETLDLQ-----NSQLPOLFPDTPRGAPAGSVKQAIINDL 843
Db 724 FEPSPYPTT--SSLEDFVTCQLPKNQHGILPQSAITTPQI--CYAGAV-----SM 771
Qy 844 MOLTAE-----NSPVTPVGAOKTALRISOSTFNNPFGOLGRLLPQNLPLDITL 893
Db 772 YQCGPEPQHTHWGQMXYNPLP--GQAFALNKFQNGVNLMEYTPAELENNINNTQTTHQLPL 830
Qy 894 OSPTGAGFPFPIRNS 908
Db 831 HHPSEARFPFDLTSS 845

```

Thu Sep 12 10:53:00 2002

us-09-842-256-2.ra1

Page 32

Search completed: September 7, 2002, 10:38:25  
Job time: 287 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 12, 2002, 03:35:41 : Search time 6106.83 Seconds  
(without alignments)  
13605.628 Million cell updates/sec

Title: US-09-842-256-1

Perfect score: 6156  
Sequence: 1 GCGCGCCGACGCTCGGCTA.....CCACTTTATGTTGTTTTT 6156

Scoring table:

IDENTITY\_NUC  
Gapop 10.0, Gapept 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

EST.\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estnu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hlc:\*  
9: gb\_estl:\*  
10: gb\_est2:\*  
11: gb\_hlc:\*  
12: gb\_gss:\*  
13: em\_gss\_hum:\*  
14: em\_gss\_inv:\*  
15: em\_gss\_pln:\*  
16: em\_gss\_vrt:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	839.2	13.6	1003	10	BM459228
2	746.8	12.1	918	10	BM451917
3	698	11.3	891	10	BM451917
4	674.2	11.0	864	10	BM451917
5	639.4	10.4	1070	10	BM451917
6	634	10.3	742	9	BM451917
7	553	9.0	716	10	BM451917
8	548.6	8.9	618	9	BM451917
9	543.8	8.8	941	10	BM451917
10	534.2	8.7	610	10	BM451917
11	520.8	8.5	524	9	BM451917
12	519.6	8.4	555	9	BM451917
13	493.2	8.0	631	9	BM451917
14	492.4	8.0	566	10	BM451917
15	481.2	8.0	545	9	BM451917
16	487.6	7.9	543	9	BM451917
17	483.6	7.9	655	9	BM451917

18	483	7.8	559	10	BG835556
19	480.8	7.8	542	9	AM465013
20	480	7.8	484	9	AM298513
21	478.2	7.8	555	10	BM090425
22	473.6	7.7	777	10	BG172538
23	464.6	7.5	592	10	BG990265
24	462.2	7.5	467	9	AA744255
25	446.8	7.3	466	9	AA504723
26	445.2	7.2	906	10	BF168417
27	444.2	7.2	556	9	AV666871
28	443	7.2	629	9	BB658506
29	439.4	7.1	573	10	BG923396
30	437.2	7.1	664	9	BE170504
31	434.2	7.1	688	9	BB655333
32	433.4	7.0	435	9	AT127080
33	424	6.9	537	10	BF350947
34	423.8	6.9	440	9	AT654689
35	423.2	6.9	580	9	AM258528
36	422.2	6.9	553	10	BF418851
37	420.8	6.8	448	9	AM027474
38	413	6.7	413	9	AT393449
39	408.8	6.6	485	10	BF651680
40	399.6	6.5	410	9	AT698694
41	397.6	6.5	619	9	AL596748
42	396.8	6.4	401	9	AA994479
43	395.8	6.4	665	9	BB613358
44	382.4	6.2	384	9	AU185018
45	378	6.1	681	9	BB640117

#### ALIGNMENTS

RESULT 1  
LOCUS BM459228 1003 bp mRNA linear EST 05-FEB-2002  
DEFINITION AGENCOURT 6415526 NIH\_MGC\_85 Homo sapiens cDNA clone IMAGE:5495761  
5', mRNA sequence.  
ACCESSION BM459228  
VERSION BM459228.1 GI:18508268  
SOURCE EST.  
ORGANISM human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 1003)  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs@mail.nih.gov  
Tissue Procurement: Lou Staudt  
CDNA Library Preparation: Life Technologies, Inc.  
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:  
http://image.llnl.gov  
Plate: LLM12123 row: d column: 02  
High quality sequence stop: 700.  
Location/Qualifiers  
1..1003  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5495761"  
/clone\_id="NIH\_MGC\_85"  
/tissue\_type="lymphoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: lymph. Vector: pCMV-Sport0; Site:1: NotI; Site:2: SalI; Cloned unidirectionally; Oligo-dT primed. Average insert size 1.867 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH\_MGC Library."

BASE COUNT	289 a	298 c	226 g	178 t	2 others
ORIGIN					
Query Match	13.6%;	Score 839.2;	DB 10;	Length 1003;	
Best Local Similarity	94.1%;	Pred. No. 4.5e-187;			
Matches 871;	Conservative	0;	Mismatches 55;	Indels 0;	Gaps 0;
OY 3494	CAGTGATCCAAACAGTCTCAAGTCAGAGATTCCAACATCATGCTGGAGCAAGAGGGC	3553			
Db 1	CAGTGATCCAAACAGTCTCAAGTCAGAGATTCCAACATCATGCTGGAGCAAGAGGGC	60			
OY 3554	CCGTTTCCACAGCAGTATGCATCTCAGCACAATATGGCCAGGGTAGCTATTTCGCCA	3613			
Db 61	CCGTTTCCACAGCAGTATGCATCTCAGCACAATATGGCCAGGGTAGCTATTTCGCCA	120			
OY 3614	TGCAGATCCAAACTTTCACACCATGAGGACAGCGGCTGTATATGACACATCCGATATG	3673			
Db 121	TGCAGATCCAAACTTTCACACCATGAGGACAGCGGCTGTATATGACACATCCGATATG	180			
OY 3674	AGCCAGACGGGGCTCAGAGCCACAGGGGCTATGTGAGAACAGCCAAATCAACTAAGAC	3733			
Db 181	AGCCAGACGGGGCTCAGAGCCACAGGGGCTATGTGAGAACAGCCAAATCAACTAAGAC	240			
OY 3734	TTCAACTTCAGATCGCTCCCAAGCACAGCAGATGCGCAGCCTTATGAATCAATTA	3793			
Db 241	TTCAACTTCAGATCGCTCCCAAGCACAGCAGATGCGCAGCCTTATGAATCAATTA	300			
OY 3794	GGAATGTTCCAAATGSACTTGACCTGAGGCTGTAGATACCAACAGGCACTTATTA	3853			
Db 301	GGAATGTTCCAAATGSACTTGACCTGAGGCTGTAGATACCAACAGGCACTTATTA	360			
OY 3854	ATGCACAGATGCTGGGCCAGAGACAGAGGAAATCCTGAACAGCATCTTCGACAGAGAC	3913			
Db 361	ATGCACAGATGCTGGGCCAGAGACAGAGGAAATCCTGAACAGCATCTTCGACAGAGAC	420			
OY 3914	AAATGATAGCAACACAGCAAGTTTCAGCAACGAATTTGATGAGAGACAAAGGTTGA	3973			
Db 421	AAATGATAGCAACACAGCAAGTTTCAGCAACGAATTTGATGAGAGACAAAGGTTGA	480			
OY 3974	ATATGACACCAAGCATGTGGCTCCATGATGATGACACAACTATGAGCAACCTTCGGA	4033			
Db 481	ATATGACACCAAGCATGTGGCTCCATGATGATGACACAACTATGAGCAACCTTCGGA	540			
OY 4034	TTTCCCAGCAAAATGCACAGCAGATTTCCATTTCTCTCCAAACTATACGAAATATGACAAAC	4093			
Db 541	TTTCCCAGCAAAATGCACAGCAGATTTCCATTTCTCTCCAAACTATACGAAATATGACAAAC	600			
OY 4094	CTGATCCAGGCTTACTGTGGGCTAGACATCCCAAGGCCACTTATGTACCCCGAATGG	4153			
Db 601	CTGATCCAGGCTTACTGTGGGCTAGACATCCCAAGGCCACTTATGTACCCCGAATGG	660			
OY 4154	CACATACACAGAGTCCCATGATGCACAAGTCTCAGGCCAACCCAGCCTATCAGGCCCTT	4213			
Db 661	CACATACACAGAGTCCCATGATGCACAAGTCTCAGGCCAACCCAGCCTATCAGGCCCTT	720			
OY 4214	CCGACATTAATGGATGGGGCAGAGGGAATATGGGCGGAAACAGCATGTTTTCCACAGCT	4273			
Db 721	CCGACATTAATGGATGGGGCAGAGGGAATATGGGCGGAAACAGCATGTTTTCCACAGCT	780			
OY 4274	CCCCACACACTTTTGGGACAGCAAGCAAAACAGCAGATGTACAGTAAACATGAACATGA	4333			
Db 781	CCCCACACACTTTTGGGACAGCAAGCAAAACAGCAGATGTACAGTAAACATGAACATGA	840			
OY 4334	ATGTGTCATGCGGACCAACACAGGTGCATGACAGCATGAACCAAGATGACAGACAGA	4393			
Db 841	ATGTGTCATGCGGACCAACCAACAGGTGCATGACAGCATGAACCAAGATGACAGACAG	900			
OY 4394	TCAGCATGACCTTCAGTACCTCGTG 4419				
Db 901	GGACAGGATCAGATGGACCTCAGTG 926				

[illegible]



```
|||||
Db 361 TGCCAGCGGTTCTGTAGACAGACAGCCGCGCTGTTCCAGGGTGGCGTCACT 420
Qy 4717 CGGCTGGGCGAGGAGGAGCTCCCTCTCTGACAGCTGCAAGTCCGATCCAGACGT 4776
Db 421 CGGCTGGGCGAGGAGGAGCTCCCTCTCTGACAGCTGCAAGTCCGATCCAGACGT 480
Qy 4777 CGCTCAGCTCTGCTCCGATTCACCTAGTGAACCTAGATCTCTCTCCCAAGTAA 4836
Db 481 CGCTCAGCTCTGCTCCGATTCACCTAGTGAACCTAGATCTCTCTCCCAAGTAA 540
Qy 4837 TGTGACAGGCCAATTCATACCCATGTCAGATGTAATGTAATGTAATGTAAT 4896
Db 541 TGTGACAGGCCAATTCATACCCATGTCAGATGTAATGTAATGTAATGTAAT 600
Qy 4897 GGAACACCATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4956
Db 601 GGAACACCATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660
Qy 4957 TTCCCTGGCTAACAGTCTAGTCCAAAGATTAAGATTTATCTGGGGGAAAGAAAGAT 5016
Db 661 TTCCCTGGCTAACAGTCTAGTCCAAAGATTAAGATTTATCTGGGGGAAAGAAAGAT 720
Qy 5017 TTTTAAAAAAATTAAGATGTTTAAAGCTAAAGCCTGAATTTGGATGGAAGCAG 5076
Db 721 TTTTAAAAAAATTAAGATGTTTAAAGCTAAAGCCTGAATTTGGATGGAAGCAG 779
Qy 5077 GACAGACA-CCGTGACACAGCCCTGTATTTACAGACACCCAGTGGCTGAAGACCAACA 5135
Db 780 GACAGACACCCGTGACACAGCCGTATTTACAGACACCCAGTGGCTGAAGACCAACA 839
Qy 5136 AGTCACAGTC 5145
Db 840 ACAAAAGTTC 849

RESULT 3
BI462410 891 bp mRNA linear EST 21-AUG-2001
LOCUS 603203776F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:5269674 5,
DEFINITION mRNA sequence.
ACCESSION BI462410
VERSION BI462410.1 GI:15253066
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 891)
AUTHORS NIH-MGC http://mgc.ncl.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: c9abs-remail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shitaki
Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov
Plate: LHAM1680 row: g column: 19
High quality sequence stop: 744.

FEATURES
source
location/Qualifiers
1..891
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5269674"
/clone_lib="NIH_MGC_97"
/lab_host="DH10B"
/note="Organ: testis; Vector: pBluescriptR (modified
pBluescript KS+); Site_1: BamHI; Site_2: SalI-xhoI (gtcgag
```

```
); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTTTVN-3',
size selected for average insert size 2.2 kb and
normalized to ROP 5. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NHGRI/NHGR1, National
Institutes of Health). Note: this is a NIH_MGC Library."

BASE COUNT 283 a 168 c 175 g 265 t
ORIGIN

Query Match 11.3%; Score 698; DB 10; Length 891;
Best Local Similarity 94.7%; Pred. No. 9.8e-154;
Matches 788; Conservative 0; Mismatches 35; Indels 9; Gaps 6;

Qy 4984 GATTAGATTTTATCTGGGGGAAAGAAATTTTAAAAAATTAAGTAAGATGTT 5043
Db 5 GATTAGATTTTATCTGGGGGAAAGAAATTTTAAAAAATTAAGTAAGATGTT 64
Qy 5044 TTAAGCTAAAGCCTGAATTTGGGATGGAAGAGACAGACCCGAGACGCTGATTT 5103
Db 65 TTAAGCTAAAGCCTGAATTTGGGATGGAAGAGACAGACCCGAGACGCTGATTT 124
Qy 5104 TACAGACACACCGAGTGGCTGAAGACCAACAAAGTCAAGTCTATCTAGAAAGCTCT 5163
Db 125 TACAGACACACCGAGTGGCTGAAGACCAACAAAGTCAAGTCTATCTAGAAAGCTCT 184
Qy 5164 AAAGACCATGTTGGAAAGAGTCTCCAGTCTACTGAACAGATGAAGAGAGCTGTGAGAG 5223
Db 185 AAAGACCATGTCGGAAGAGTCTCCAGTCTACTGAACAGATGAAGAGAGCTGTGAGAG 244
Qy 5224 GCTGTTAATTAAGTAATTTTCTCTGTTTCTTTCTTTTCTTTTCTTTTCTTTTCTTTT 5283
Db 245 GCTGTTAATTAAGTAATTTTCTCTGTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 304
Qy 5284 ACCTGAATCATGAATTAAGTAATTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 5343
Db 305 ACCTGAATCATGAATTAAGTAATTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 363
Qy 5344 GATCAGACAGCTTGAATCAGATCTCTCTGTCAGGCTGACGCTCCCTCCCT 5403
Db 364 GATCAGACAGCTTGAATCAGATCTCTCTGTCAGGCTGACGCTCCCTCCCT 423
Qy 5404 CTCTCATTTCCCATACCTCCCTATTTTCTCTCTTTTAAAAAATTAATTAAGTACAGA 5463
Db 424 CTCTCATTTCCCATACCTCCCTATTTTCTCTCTTTTAAAAAATTAATTAAGTACAGA 483
Qy 5464 AACGAGTAAAGCCCTTATTTCTTTAAATGTTTGGCCAGCCACTTACCAATTTGTAAGTA 5523
Db 484 AACGAGTAAAGCCCTTATTTCTTTAAATGTTTGGCCAGCCACTTACCAATTTGTAAGTA 543
Qy 5524 TTGAATTTCAAGAAAAAATTAAGTATTTCTGCAAGAGAGCAAAAGTTAAGCT-T 5582
Db 544 TCGAATTTCAAGAAAAAATTAAGTATTTCTGCAAGAGAGCAAAAGTTAAGGCTGT 603
Qy 5583 GATACCAATGACGTAAGATACCTGCTTTGGAAGCATGTTATTTCTTCCCAAGCAAC 5642
Db 604 GATACCAATGACGTAAGATACCTGCTTTGGAAGCATGTTATTTCTTCCCAAGCAAC 663
Qy 5643 TCTGAGCTCCAAATGAGAG-AAAGCCAGAGTCTTTAA-ATTGATGACAGATTTACAGA 5700
Db 664 TCTGAGCTCCAAATGAGAGAAAGCCAGAGTCTTTCAACATGATAGACATATACAGA 723
Qy 5701 CAGATTTAACCTGTCATGTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 5757
Db 724 CAGATTT-AACTCTGATGTTGTTTCAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 781
Qy 5758 CGAAGTTTGTGAAGTACATTAATTTCAATTTATATGTAAACAGCAATAT 5809
Db 782 CGAAGTTTGTGAAGGACATGAATTTCAATTTATTTTGTCAACAGCAAT 833

RESULT 4
```

```

BI734567      BI734567      864 bp      mRNA      linear      EST 20-SEP-2001
LOCUS         603356396F1 NIH_MGC_94 Mus musculus cDNA clone IMAGE:5363950 5',
DEFINITION    mRNA sequence.
ACCESSION     BI734567
VERSION       BI734567.1 GI:15711580
KEYWORDS      EST.
SOURCE        house mouse.
ORGANISM      Mus musculus
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE     1 (bases 1 to 864)
AUTHORS       NIH-MGC http://mgc.nci.nih.gov/.
TITLE         National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL       Unpublished (1999)
COMMENT       Contact: Robert Strausberg, Ph.D.
               Email: cgraphs-remail.nih.gov
               Tissue Procurement: The Cepko Laboratory
               cDNA Library Preparation: Life Technologies, Inc.
               DNA Sequencing by: Incyte Genomics, Inc.
               Clone distribution: MGC clone distribution information can be
               found through the I.M.A.G.E. Consortium/LLNL at:
               http://image.llnl.gov
               Plate: LHAM11925 row: 0 column: 23
               High quality sequence stop: 828.
               Location/Qualifiers
                 1..864
                 /organism="Mus musculus"
                 /db_xref="taxon:10090"
                 /clone_1fb="IMAGE:5363950"
                 /clone_1lb="NIH_MGC_94"
                 /tissue_type="retina"
                 /lab_host="DH10B (phage-resistant)"
                 /note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
                 Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                 Average insert size 3.3 Kb. Library enriched for
                 full-length clones and constructed by Life Technologies.
                 Note: this is a NIH-MGC Library."

BASE COUNT    223 a      257 c      225 g      159 t
ORIGIN
Query Match    11.0%; Score 674.2; DB 10; Length 864;
Best Local Similarity 89.3%; Pred. No. 4.1e-148;
Matches 771; Conservative 0; Mismatches 88; Indels 4; Gaps 4;

QY 1403 GTGAGACATGACCTCAGACATATATATTTCCCATTAATGGCCCAAGAACAA 1462
    |||||||
Db 1 GTGAGACATGACCTCAGACATATATATTTCCCATGAATGGCCCAAGAACAA 60
    |||||||

QY 1463 TGGGCAATGCCATGGCAGGTTGTGTTCTGGGGAAATGAACATGTCAGGCATGC 1522
    |||||||
Db 61 TGGGCAATGCTATGGCAGGTTGTGTTCTGGGGCAATGAACATGTCAGGCATGC 120
    |||||||

QY 1523 AAGCAACCACTCTCTCAGGGTAGTACTAGTCAAAATGAACAGCCCTCAACAGCA 1582
    |||||||
Db 121 AAGCAACCACTCTCTCAGGGTAGTACTAGTCAAAATGAACAGCTCTCGCAAGCA 180
    |||||||

QY 1583 GCCCTGCATGAATCCAGACACCCACCTCATGCTTTCACCAAGCATCGCATGACC 1642
    |||||||
Db 181 GCCCTGCATGAACCCGGGGCAAGCCAGCTCGCTCTCCCAAGGACAGCATGAGCC 240
    |||||||

QY 1643 CTGAGATGGCTGGCAGCCTCGAATCCACCCAGTCAGTTTCCCTCGAGGAAGCTGC 1702
    |||||||
Db 241 CCGGCGTGGTGGCAGTCCCTCGCATCCACCCAGTCAGTTTCCCTCGGAGGAGCTTGC 300
    |||||||

QY 1703 ATTCCCTGTGGAGTTTGGCAGCAGCAGAGAAATAGCCATAGTTATACCAACAGCTCCC 1762
    |||||||
Db 301 ATTCCCTGTGGAGTTTGGCAGCAGCAGAGAAATAGCCATAGTTATACCAACAGTTCCC 360
    |||||||

QY 1763 TCATATGACTTCAGGCCCTCAGCGAGGGGCGAGGGGTCTATTAGGGTCAATCGTTGGCTT 1822
    |||||||

```

---

```

Db 361 TCATATGACTTCAGGCCCTCAGCGAGGGGCGAGGGGTCTACTCGGGTCTCGTGGCTT 420
QY 1823 CACCAACCTTAATAATGGCAATTTTGCATAAAGCTCCCACTTATATGAATCTCCCCAC 1882
    |||||||
Db 421 CACCGACCTAATAATGGCAATTTTGCATAAAGCTCCCACTTATATGAATCTCCCCAC 480
    |||||||

QY 1883 TCAGCAAGATGGGAGCTTGGAGCTCAAAAAGACTGTTTGGACTATATGGGAGCCCTCAG 1942
    |||||||
Db 481 TCAGCAAGATGGGAGCTTGGAGCTCAAAAAGACTGTTTGGACTATATGGGAGCCCTCAG 540
    |||||||

QY 1943 AAGGTACACTGGACAGACAGAGAGAGAGCTGCATCTGAGAGCAAAAGAAACAAATG 2002
    |||||||
Db 541 AAGGTACAACTGGACAGACAGAGAGAGAGCTGCATCTGAGAGCAAAAGAGGAGGAGGAG 600
    |||||||

QY 2003 ACCCCACCTGCCCCCGCCGCTGAGCAGTGAAGAGAGCTGACGGGACAGACAGCTGCATG 2062
    |||||||
Db 601 ATTCCAGCATGCCCCAGGCGGCGAGGAGACAGGCTGAGGAGACACAGCCGCTGCATG 660
    |||||||

QY 2063 ACAGCAAGGGCAGACCAAACTCTGACCTGCTGACCAACCAATCTGA-TGAGATGAG 2121
    |||||||
Db 661 ACAGCAAGGGCAGACCAAACTCTGACCTGCTGACCAACCAAGTCCGACCAAGATGGAG 720
    |||||||

QY 2122 CCCTGGCCCTTAG-CCAGCTCTTGTGGATAC-AAACAAGACTCCACAGGATGCTGC 2179
    |||||||
Db 721 CCTTACCTCTTGGCCAGCTCTTGTGGACACAAAACAGGACTCAACAGGAGAGCTTGC 780
    |||||||

QY 2180 CTGGTTC-TGGGTCATACATGAGAACCTGCTCAAGAGAGAGCAATTTTGCACAGA 2238
    |||||||
Db 781 CTGGGCTTTGGGTCACAGCATGTCGCTGCTCAAGAGAGAGATTAAGATTTTGCACAGA 840
    |||||||

QY 2239 CTCTTGACAGACAGCATGTTCCC 2261
    |||||||
Db 841 CTCTTACGAGCAAGCATGTTCCC 863
    |||||||

RESULT 5
BI248968      BI248968      1070 bp      mRNA      linear      EST 17-JUL-2001
LOCUS         602992572F1 NCI_CGAP_Mam5 Mus musculus cDNA clone IMAGE:5148545 5',
DEFINITION    mRNA sequence.
ACCESSION     BI248968
VERSION       BI248968.1 GI:14795876
KEYWORDS      EST.
SOURCE        house mouse.
ORGANISM      Mus musculus
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE     NIH-MGC http://mgc.nci.nih.gov/.
AUTHORS       National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE         Unpublished (1999)
JOURNAL       Contact: Robert Strausberg, Ph.D.
               Email: cgraphs-remail.nih.gov
               Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys
               cDNA Library Preparation: Life Technologies, Inc.
               DNA Sequencing by: Incyte Genomics, Inc.
               Clone distribution: MGC clone distribution information can be
               found through the I.M.A.G.E. Consortium/LLNL at:
               http://image.llnl.gov
               Plate: LHAM11366 row: p column: 18
               High quality sequence stop: 905.
               Location/Qualifiers
                 1..1070
                 /organism="Mus musculus"
                 /strain="C57/B6"
                 /db_xref="taxon:10090"
                 /clone_1fb="IMAGE:5148545"
                 /clone_1lb="NCI CGAP_Mam5"
                 /tissue_type="tumor, gross tissue"
                 /dev_stage="7 months"
                 /lab_host="DH10B"
                 /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;

```



QY 4945 TCTTGGCTTTGTTTCCCTGGCTAACAGTCTAGTCCCAAGATTATATCTGGGG 5004  
 Db 361 TCTTGGCTTTGTTTCCCTGGCTAACAGTCTAGTCCCAAGATTATATCTGGGG 420  
 QY 5005 AAGAGAAAAGATTTTAAAAATTAAGTATTAAGCTTAAGCCCTGAATTGG 5064  
 Db 421 AAGAGAAAAGATTTTAAAAATTAAGTATTAAGCTTAAGCCCTGAATTGG 480  
 QY 5065 GGATGAGACAGACAGACACCGCTGATATTATACAGACACCCAGTCCGTG 5124  
 Db 481 GGATGAGACAGACAGACACCGCTGATATTATACAGACACCCAGTCCGTG 540  
 QY 5125 AAGACCAACAAAGTCAAGCTGATCTAGAAAAGCTTAAGACCATGTTGGAAGAGT 5184  
 Db 541 AAGACCAACAAAGTCAAGCTGATCTAGAAAAGCTTAAGACCATGTTGGAAGAGT 600  
 QY 5185 CTCACATTTACTGACAGATGAAGAAGAGCTGAGAGAGGCTTTAATACCAATAT 5244  
 Db 601 CTCACATTTACTGACAGATGAAGAAGAGCTGAGAGAGGCTTTAATACCAATAT 660  
 QY 5245 TTTTTCCTTGTCTTCTT 5263  
 Db 661 TTTTTCCTTGTCTTCTT 679

## RESULT 7

BI156653 716 bp mRNA linear EST 05-JUL-2001  
 LOCUS 60292125F1 NIH\_CGAP\_Mam3 Mus musculus cDNA clone IMAGE:5061685 5',  
 DEFINITION mRNA sequence.

ACCESSION BI156653  
 VERSION BI156653.1 GI:14616654  
 KEYWORDS EST.

SOURCE house mouse.  
 ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 716)  
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-r@mail.nih.gov

Tissue Procurement: Lohar Hennighausen Ph.D., Chu-Xia Deng Ph.D.  
 cDNA Library Preparation: Life Technologies, Inc.  
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: LAM11166 row: m column: 14

High quality sequence stop: 714.

## FEATURES

Location/Qualifiers  
 1..716

/organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:5061685"  
 /clone\_lib="NIH\_CGAP\_Mam3"  
 /tissue\_type="tumor, gross tissue"  
 /lab\_host="DH10B"

/note="Organ: mammary; Vector: pCMV-SPORT6; Site\_1: NotI;  
 Site\_2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
 Average insert 2 kb. Library constructed by Life  
 Technologies, catalog #12017-018. Investigators providing  
 samples: Lohar Hennighausen/Chu-Xia Deng, NIH Reference  
 for transgenic model: Xu et al., Nature Genetics 22, 37-43  
 (1999). Note: this is a NCI/CGAP Library."

BASE COUNT 189 a 222 c 177 g 128 t

## ORIGIN

Query Match 9.0%; Score 553; DB 10; Length 716;  
 Best Local Similarity 87.3%; Pred. No. 1.7e-119;

Matches 631; Conservative 0; Mismatches 85; Indels 7; Gaps 2;  
 QY 3727 CTAGACCTTCAACTTGAAGTATGAGTCCCTCAGGACAGACAGATATGCCAGCATATGAAT 3786  
 Db 1 CTAGACCTTCAACTTGAAGTATGAGTCCCTCAGGACAGACAGATATGCCAGCATATGAAT 60  
 QY 3787 CAAATCAGCAATGTTTCCAAATGGAATGAACTTGAAGCTTGAAGCTTGAATGAGAGACAA 3846  
 Db 61 CAAATCAGCAATGTTTCCAAATGGAATGAACTTGAAGCTTGAAGCTTGAATGAGAGACAA 120  
 QY 3847 CCTATTAAATGACAGATGCTGGGCCAGAGACAGAGGAAATCTGAGATATCTGCA 3906  
 Db 121 CCTATTAAATGACAGATGCTGGGCCAGAGACAGAGGAAATCTGAGATATCTGCA 180  
 QY 3907 CAGAGCAAAATGATGATGACAGCAACAGCAAGTTTCAAGCACTTGAATGAGAGACAA 3966  
 Db 181 CAGAGCAAGT-----GCAGCAGAGGTGACAGACGGACTGTGATGATGAGAGACAG 234  
 QY 3967 GGGTGAATATGACACCAAGCATGTGGCTCTTAATGATGACCAACATATGACCAAC 4026  
 Db 235 GGGTGAATATGACACCAAGCATGTGGCTCTTAATGATGACCAACATATGACCAAC 294  
 QY 4027 CCTGGATTTCCCGAGCAAAATGACAGAGTTCCTTCCATTCCTCCAAATGAGGAATAGT 4086  
 Db 295 CCTGGATTTCCCGAGCAAAATGACAGAGTTCCTTCCATTCCTCCAAATGAGGAATAGT 354  
 QY 4087 CAGCAACCTGATCAGAGCTTCTTACTGAGGCTACAGACTCCCGAGAGCCATATATGACCC 4146  
 Db 355 CAGCAACCTGATCAGAGCTTCTTACTGAGGCTACAGACTCCCGAGAGCTCTTAATGATCCG 414  
 QY 4147 CGAATGACACATATACAGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 4206  
 Db 415 CGAATGACACATATCAGAGTCCATGATGATGATGATGATGATGATGATGATGATGATGAT 474  
 QY 4207 GCGGCTCCGAGCATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4266  
 Db 475 GCGGCTCCGAGCATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 534  
 QY 4267 CAGCAGTCCCGACACATGTTGGGCGAGCAAGCAACAGCATATGATGATGATGATGATGAT 4326  
 Db 535 CAGCAGTCCCGACACATGTTGGGCGAGCAAGCAACAGCATATGATGATGATGATGATGAT 594  
 QY 4327 AACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4386  
 Db 595 AACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 654  
 QY 4387 GGACATATGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4446  
 Db 655 GGACATATGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 713  
 QY 4447 CCC 4449  
 Db 714 CCC 716

## RESULT 8

AW854255 618 bp mRNA linear EST 19-MAY-2000  
 LOCUS RC3-CT0254-100500-211-d11 CT0254 Homo sapiens cDNA, mRNA sequence.  
 DEFINITION  
 ACCESSION AW854255  
 VERSION AW854255.1 GI:7949948  
 KEYWORDS EST.

SOURCE human.

## ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 618)

AUTHORS Dias Neto,E., Garcia Correa,R., Verjowski-Almeida,S., Briones,M.R.,  
 Nagai,M.A., da Silva,M. Jr., Zago,M.A., Bordin,S., Costa,F.F.,  
 Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,  
 Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,  
 M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and  
 Simpson,A.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
MEDLINE 20202663  
COMMENT Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Genome  
Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?ti=6t2-RC3-CT0254-100  
500-211-d11e13-2000-05-10&t4=1)  
Seq primer: puc 18 forward  
High quality sequence start: 70  
High quality sequence stop: 144.  
Location/Qualifiers  
1. 618

FEATURES  
source  
1. 618  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_id="CT0254"  
/dev\_stage="Adult"

/note="Organ: colon; Vector: puc18; Site.1: Sma1; Site.2:  
Sma1; A mini-library was made by cloning products derived  
from ORESTES PCR (U.S. Letters Patent application No. 196  
, 716 - Ludwig Institute for Cancer Research) profiles  
into the puc 18 vector. Reverse transcription of tissue  
mRNA and cDNA amplification were performed under low  
stringency conditions."

BASE COUNT 199 a 116 c 121 g 182 t  
ORIGIN

Query Match 8.9%; Score 548.6; DB 9; Length 618;  
Best Local Similarity 96.4%; Pred. No. 1.8e-118;  
Matches 593; Conservative 0; Mismatches 19; Indels 3; Gaps 3;

QY 4750 ACAGCTGAAAGCTGCATCCAGACAGTGCCTGCTGCTCCGATTCACCTTAACTGC 4809  
DB 5 ACAGGCTGATGCTCGCAT-CAGACAGTGCCTGCTGCTCCGATTCACCTTAACTGC 63  
QY 4810 AACTTATATCTCTCTCCCAAGTAATGTGACAGCCATTTCAATCCATGTCAGAT 4869  
DB 64 AACTTATATCTCTCTCCCAAGTAATGTGACAGCCATTTCAATCCATGTCAGAT 123  
QY 4870 TGAATGATTTAATGATATGATTTAAG-GAAGACCATGCTCTTGTCTGTTCTGTTG 4928  
DB 124 TGAATGATTTAATGATATGATATGATTAAGAGAACATGCTCTTGTCTGTTCTGTTG 183  
QY 4929 GTTCCAGACACTGCTTCTGCTTGTTCCTGCTGCTAAGAGTCTAGTGCCTAAGATTA 4988  
DB 184 GTTCCAGACACTGCTTCTGCTTGTTCCTGCTGCTAAGAGTCTAGTGCCTAAGATTA 243  
QY 4989 AGATTTATATCGGG-GGAAGAGAAAGAAATTTTAAATAAATAAAGATGTTTAA 5047  
DB 244 AGATTTATATCGGGAGAGAGAAAGAAATTTTAAATAAATAAAGATGTTTAA 303  
QY 5048 GCTAAACCTGAATTTGGATGAGAGACAGACACCTGAGACGCTGATTTTACA 5107  
DB 304 GCTAAACCTGAATTTGGATGAGAGACAGACACCTGAGACGCTGATTTTACA 363  
QY 5108 GACACACCCAGTGGGTGAAGACCAAGTACAGTGTATCTGTAGAAAGCTCTAAG 5167  
DB 364 GACACACCCAGTGGGTGAAGACCAAGTACAGTGTATCTGTAGAAAGCTCTAAG 423  
QY 5168 ACCATGTTGAGAAAGTCTCCAGTTTACGAGACAGATGAAAGAGAGCCGTAGAGGGGTG 5227  
DB 424 ACCATGTTGAGAAAGTCTCCAGTTTACGAGACAGATGAAAGAGAGCCGTAGAGGGGTG 483  
QY 5228 TTAACATTAAGCAATATTTTCTTCTGTTTAAACCAAACTGTGTTACCT 5287

DB 484 TTAACATTAAGCAATATTTATCTCTGTATTAACCTGTGTTAAACCAAACTGTGTTACCT 543  
QY 5288 GAATCATGAATTTGAGAGAGAAATTAATTTCAATTTCAATTAAGTCCCTTTAGTTGATC 5347  
DB 544 GAATCATGAATTTGAGAGAGAAATTAATTTCAATTTCAATTAAGTCCCTTTAGTTGATC 603  
QY 5348 AGACAGCTGAATCA 5362  
DB 604 AGACAGCTGAATCA 618

RESULT 9  
B1694248 941 bp mRNA linear EST 18-SEP-2001  
LOCUS 603347586F1 NCL\_CGAP\_Mam2 Mus musculus cDNA clone IMAGE:5375468 5,  
DEFINITION mRNA sequence.  
ACCESSION B1694248  
VERSION B1694248.1 GI:15656877  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 941)  
NTH-MGC http://mgc.ncl.nih.gov/.  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: c9abs-remail.nih.gov  
Tissue Procurement: Gilbert Smith, Ph.D.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LMNL at:  
http://image.llnl.gov  
Plate: LMNL955 row: 0 column: 21  
High quality sequence stop: 800.  
Location/Qualifiers  
1. 941

FEATURES  
source  
1. 941  
/organism="Mus musculus"  
/strain="FVB/N-3"  
/db\_xref="taxon:10090"  
/clone\_image="5375468"  
/clone\_id="NCL\_CGAP\_Mam2"  
/tissue\_type="tumor, biopsy sample"  
/dev\_stage="5 months"  
/lab\_host="DH10B"  
/note="Organ: mammary; Vector: pCMV-SPORT6; Site.1: SalI;  
Site.2: NotI; Cloned unidirectionally. Primer: Oligo dT.  
Library constructed by Life Technologies. Investigator  
providing samples: Gilbert Smith, NTH"

BASE COUNT 244 a 265 c 251 g 181 t  
ORIGIN

Query Match 8.8%; Score 543.8; DB 10; Length 941;  
Best Local Similarity 80.0%; Pred. No. 2.6e-117;  
Matches 730; Conservative 0; Mismatches 167; Indels 16; Gaps 7;

QY 3780 TATGAATCAATAGCAAGATGTTCCAAATGCACTGACTGAGGCTGAGTACCAAC 3839  
DB 1 TATGAATCAATAGCAAGATGTTCCAAATGCACTGACTGAGGCTGAGTACCAAC 58  
QY 3840 ACAGGACCTATTAATGACAGATGCTGGCCAGAGACAGAGGAGAAATCTGAACCAACA 3899  
DB 59 TCAGGCTCCTATTAATGACAGATGCTGGCCAGAGACAGAGGAGAAATCTGAACCAACA 118  
QY 3900 TCTTTCAGACAGCAATATGATCAAGCAAGCAAGTTCAGCAAGCAATTTGATGATGAG 3959  
DB 119 CCTTCGGCAGAGCGAGAT-----GCAGCAGCAGAGTGCAGCAGCGAGCTGATGATGAG 172

```

Oy 3960 AGGACAGGGTTGATATATACACCAAGCATGTGGCTCCTAGTGTATGCCAGCACTAT 4019
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 173 AGGACAGGGCTTGAATGATACCCAGCATGTGGCTCCCGCTCCAGCAGCAGCAT 232
Oy 4020 GAGCAACCCCTGGATCCCGAGGAAATGCACAGCTTTCATTCCTCCAACTAGCG 4079
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 233 GAGCAATCCCGGATCCCGAGGCAATGCGCCAGCTTCCCTTCCTCCGAACTAGCG 292
Oy 4080 AATTAAGTCAGCAACTGATCCAGGCTTACTGGGCTAGCACTCCAGAGCCCACTTAT 4139
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 293 AATTAAGTCAGCAACTGATCCAGGCTTACTGGGCTAGCACTCCAGAGCTTCTTAT 352
Oy 4140 GTACCCGGAATGCAATACACAGAGTCCCATGATGCAACAGTCTCAGGCCAACCCAGC 4199
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 353 GTTCCCTCGGATGGCAGATACCTCAGAGTCCCATGATGACAGCTCTCAAGCCAAACCAGC 412
Oy 4200 CATATCGGCCCCCTCCGATATAATNGATGGCGCAGGGGAAACATGGGGGAAACAGCAT 4259
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 413 CTACCGCCCGCTCAGACATGATGATGGGACAGGGGAGCATGGGTGAAACAGCAT 472
Oy 4260 GTTTTCCAGAGTCCCAACCACTTTGGGAGCAAGCAAAACACAGCATGTACATTA 4319
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 473 GTTTCACAGCAGTCCCAACCACTTTGGGAGCAAGCAAAACACAGCATGTATAGTA 532
Oy 4320 CAACATGAACATCAATGTGTCCATGGGAGCAACACAGGTGGCATGAGCATGAACCA 4379
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 533 CAACATGAACATCAATGTGTGTGATGGCAACCAACAGGGGTGGCTTGAAGCATGAACCA 592
Oy 4380 GATGACAGACAGATACAGATGACCTGACATGACCTGCTCTACCTCAGGCGCTGCTC 4439
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 593 GATGACAGGCGCATGATGATGATGACCTGACATGACCTGCTCTACCTCAGGAGTGG-CCTC 651
Oy 4440 CATGGGTCCCGAGAGCATTAATGATCTGCTCTGAGGGAGGAGCAAC--TGTTCCTCAA 4496
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 652 CATGGGTCCCGAGAGCATTAATGATCTGCTCTGAGGGAGGAGCAACCTTTTCCAAAA 711
Oy 4497 CCAGCTG--CTTGGAAATGATATGATTAAGCAGAGGAGAGACAAAC-ACGGAATATTTG 4553
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 712 CCAGCTGCGCTGGAAATGATATGATCAAGCAAGAGAGATGATCATCTTCGAAATATCT 771
Oy 4554 CTGACACTCTGAGAGCATGTGCTTCTGACGTGACCGGGCTCACTTGTCTCAAAACACT 4613
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 772 GTTGACCCCGAGAAACTGTCTGATCTTCTTCAACCACTGGGTACAAAACATTTTA 831
Oy 4614 CCAGCTGAGAGA-GCTGTGCTATTTTGTTCAAACCACTGACCTGCAGCCGGTTCTGC 4672
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 832 CCAGCTGAGAGAGGCTGCTTCTGCTGTGTGGCACCCTGACATTCGCCGAGGTTTCC 891
Oy 4673 TAGAGCAGACAG 4685
    ||||| |||||
Db 892 AGGAGCATTCAG 904

```

```

RESULT 10
BF044623/c 610 bp mRNA linear EST 10-OCT-2000
LOCUS BP230020A20G6 Soares normalized bovine placenta Bos taurus cDNA
DEFINITION BP230020A20G6 5', mRNA sequence.
ACCESSION BF044623
VERSION BF044623.1 GI:10761678
KEYWORDS EST.
SOURCE
  ORGANISM Bos taurus
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
  Bovidae; Bovinae; Bos.
  1 (bases 1 to 610)
  Lewin,H.A., Soares,M.B., Rebeiz,M., Pardinaz,J., Liu,L. and Larson
  J.H.
  TITLE Bovine ESTs
  JOURNAL Unpublished (2000)
  COMMENT Contact: Lewin, H. A.
  W. M. Keck Center for Comparative and Functional Genomics

```

University of Illinois at Urbana-Champaign  
340 Edward R. Madigan Laboratory, 1201 W. Gregory Dr., Urbana, IL  
61801, USA  
Tel: 217 333 5998  
Fax: 217 244 5617  
Email: h-lewin@uiuc.edu

Funding for cattle EST sequencing was provided by the USDA National  
Research Initiative, Animal Genome Resource Grant AG 99-3205-8534  
to H. A. Lewin and J. E. Womack. Base Calling/Quality Scores: PHRED  
from Washington University Genome Center. Vector Trimmi g:  
Cross match from Washington University Genome Center PHRAP suite.  
This sequence is vector free and at least 200 bp in length.

PCR Primers  
FORWARD: TAATACGACTCATATAGG  
BACKWARD: ATTAACCTCATTAAG  
Insert Length: 610 Std Error: 0.00  
Plate: BP250020A20 row: G column: 6  
Seq primer: ACCGATTAACAATTTCACACAGGA  
High quality sequence stop: 610.

FEATURES  
source  
1..610  
/organism="Bos taurus"  
/db\_xref="taxon:9913"  
/clone="BP250020A20G6"  
/clone\_lib="Soares normalized bovine placenta"  
/sex="female"  
/lab\_host="DH10B"  
/note="Organ: Placenta; Vector: pRTT3pac; Site: 1; Ecoret;  
Site: 2: NotI. The cDNA library was contributed by the  
Soares Laboratory and it was constructed and normalized  
as described by Bonaldo, M.F., Lennon, G. and Soares,  
M.B. (1996), Genome Research 6(9): 791-806."

BASE COUNT  
132 a 156 c 132 g 186 t 4 others

ORIGIN

```

Query Match      8.7%; Score 534.2; DB 10; Length 610;
Best Local Similarity 93.0%; Pred. No. 4,4e-115;
Matches 568; Conservative 0; Mismatches 42; Indels 1; Gaps 1;

```

```

Oy 55 CCGGACCTGAGGGCGGTGAGCCGACCCGAGCGCATTTCTGTGATTTGGCTACACCTTA 114
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 610 CCGGACCTGAGGGCGGTGAGCCGACCCGAGCGCATTTCTGTGATTTGGCTACACCTTA 551
Oy 115 TAGATCTTGCAGCTGTTTACAGGACAGCTTCTGATGTGTTCAAGATGAGTGGATG 174
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 550 TAGATCTTGCAGCTGTTTACAGGACAGCTTCTGATGTGTTCAAGATGAGTGGATG 491
Oy 175 GGAGAAATACCTCTGACCCCTTCAGGGCAGAGACAGAAAGCGCAGGAATGTCTGAC 234
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 490 GGAGAAATACCCCTGACCCATTCAGGGCAGAGACCGGGAAGCGCAGGAATGTCTGAC 431
Oy 235 CAACCTGAGACCGCCGAAAGGAACACTGAAGAAAGTAATGTGAACAGGAAATTA 294
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 430 CAGCTGGACCGACCCGAAAGAGCAGCGAAGACGTACCGTGAACAGGAAATTA 371
Oy 295 TATATGAGAACTTGCAGAGTGTATTTTGCAGAAATTTTAATGATATGACAACTTTAAC 354
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 370 TATATGAGAACTTGCAGAGTGTATTTTGCAGAAATTTTAATGATATGACAACTTTAAC 311
Oy 355 TTCAAACTGACAAATGTCAATCTTAAAGAACTGTGAAGCAATTTGTGCATCA 414
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 310 TTCAAACTGACAAATGTCAATCTTAAAGAACTGTGAAGCAATTTGTGCATCA 251
Oy 415 GAACAAGAAAGACAGCAGCTGCAATGATGATGAGTGCAGAGCAATGATGCTCTCT 474
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 250 GAACAAGAAAGACAGCAGCTGCAATGATGATGAGTGCAGAGCAATGATGCTCTCT 191
Oy 475 ACAGGCGAGGCTCATTCAGCAAGAGATGCGGTGGGCTATGATGCTTAGAGCCCTTGTAT 534
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 190 ACCGGGCAAGGGGTCAATTGACAAAGATGCACTGGGCGCATATGCTGAGGCGCTTGAC 131
Oy 535 GGGTCTCTTCTTGTAGTGAACCTGGAAGCAACGTTGTGTTGTCTCAGAGAAATGTGACA 594

```

```

|||||
Db 130 GGGTCTCTTGTGCGTGAACCTGGAAGCAAGCTTGTGTTTCAGACAATG -NNNA 72
Oy 595 CAGATATCAAGAGTATTAACCAAGAGAGCTGATGAACAAAATGATATAGATCTTGAT 654
Db 71 CAGTACCTTAAGGTATTAACCAAGAGAGCTGATGAACAAAAGTGTATTAATCTCTGCAC 12
Oy 655 GTTGGGACCA 665
Db 11 GTTGGGACCA 1

RESULT 11
AA740705 524 bp mRNA linear EST 07-FEB-1998
LOCUS ny98404.s1 NCI_CGAP_GCB1 Homo sapiens CDNA clone IMAGE:1286286 3,
DEFINITION similar to TR:Q13536 Q13536 TRANSCRIPTONAL INTERMEDIARY FACTOR 2.
; mRNA sequence.
ACCESSION AA740705
VERSION AA740705.1 GI:2779297
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 524)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLES National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
JOURNAL Tumor Gene Index
COMMENT Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
Ph.D., Gerald Marti, M.D.
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/dbfp/image/image.html
Insert Length: 661 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 367.
location/Qualifiers
1.524
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1286286"
/clone_1id="NCI_CGAP_GCB1"
/tissue_type="germinal center B cell"
/lab_host="DH10B"
/note="Vector: p773D-Pac (Pharmacia) with a modified
polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from human tonsillar cells enriched for
germinal center B cells by flow sorting (CD20+, IgD-),
provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
(NCI) and Dr. Gerald Marti (CEBR). cDNA synthesis was
primed with a Not I - oligo(dT) primer
15'-TGTTACCAATCTGAAGTGGAGCGCGCTCATTTTCTTTT-3'
1. Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified p773 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT 171 a 119 c 130 g 104 t
ORIGIN

```

Query Match 8.5% Score 520.8; DB 9; Length 524;  
 Best Local Similarity 99.6%; Pred. No. 6.2e-112;  
 Matches 522; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

Oy 2147 CGATATCAAAAGAGCTCCAGAGTACCTTCCCTGGTCTGGGGTACACATGGAACCT 2206
Db 1 CGATATCAAAAGAGCTCCAGAGTACCTTCCCTGGTCTGGGGTACACATGGAACCT 60
Oy 2207 CGCTCAAGAGAGCATTAATTTTTCACAGAGCTCTTTCAGAGACAGCTTCCCTGTGG 2266
Db 61 CGCTCAAGAGAGCATTAATTTTTCACAGAGCTCTTTCAGAGACAGCTTCCCTGTGG 120
Oy 2267 ACTTGGCCAACTTAACAGACAGACAGCCAGCAAGAGCTTACAGCAGACAGCA 2326
Db 121 ACTTGGCCAACTTAACAGACAGACAGCCAGCAAGAGCTTACAGCAGACAGCA 180
Oy 2327 CAGCTCTGATCAGAGTACGATTAATTAACAAGAGCCGAGGAGCCCAAGAGAGAGAG 2386
Db 181 CAGCTCTGATCAGAGTACGATTAATTAACAAGAGCCGAGGAGCCCAAGAGAGAGAG 240
Oy 2387 ATGCACACTCTTCTGCTATTTTGTAGATTAAGATGATTAAGATATGTTTACAGAAA 2446
Db 241 ATGCACACTCTTCTGCTATTTTGTAGATTAAGATGATTAAGATATGTTTACAGAAA 300
Oy 2447 TAACCCCAAACTTGAGAGAGCTGAGACAGTAAAGACAGATCTCCAGTAAACAAATTA 2506
Db 301 TAACCCCAAACTTGAGAGAGCTGAGACAGTAAAGACAGATCTCCAGTAAACAAATTA 360
Oy 2507 TAGCAATGAAGAGAGAGAGAGAGAGAGATGAGCTTGAAGCTGAGACAGCTGAGAGT 2566
Db 361 TAGCAATGAAGAGAGAGAGAGAGAGAGATGAGCTTGAAGCTGAGACAGCTGAGAGT 420
Oy 2567 AGCTGGACAACTTGAGAGAGATTTTGTAGATTTTGCAGAAATGATTAACACAGCTTT 2626
Db 421 AGCTGGACAACTTGAGAGAGATTTTGTAGATTTTGCAGAAATGATTAACACAGCTTT 480
Oy 2627 TCCCAAGACAGAGAGAGAGAGAGAGAGATGAGCTTGAAGCTGAGAGAGAGAGT 2670
Db 481 TCCCAAGACAGAGAGAGAGAGAGAGAGATGAGCTTGAAGCTGAGAGAGAGAGT 524

RESULT 12
BE170360 555 bp mRNA linear EST 21-JUN-2000
LOCUS BE170360
DEFINITION OY4-HT0538-020300-123-e04 HT0538 Homo sapiens CDNA, mRNA sequence.
ACCESSION BE170360
VERSION BE170360.1 GI:8633081
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 555)
Dias Neto,E., Garcia Correa,R., Verjovsky-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
CONTACT: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the PAPSP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?fl=6t-QV4-HT0538-020
300-123-e04&t3=2000-03-02&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 555.

```







Best Local Similarity 88.0%; Pred. No. 2.1e-105;  
Matches 549; Conservative 0; Mismatches 73; Indels 2; Gaps 1;

```
QY 1 GGGGGCCGACCTCGGCTACAGTGTGGGGGCGCAAGTGCAGCCCGACGGCAGCCGCA 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10 GGGGGCCGACCTCGGCTACAGTGTGGGGGCGCAAGTGCAGCCCGACGGCAGCCGCA 69
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 61 CCTGAGCGGCTGACGAGCCCGGAGTTCCTTGGATTGGCTACACATTTATAGATC 120
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 70 CCTGAGCGGCTGACGAG--CCACATTGATTTTCTCGCATCTGCGCTTACATGCGCTC 127
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 121 TTCTGACCTGTTTACAGCAGAGTGTGATGATGTTTCAAGATGATGGGATGGAGAA 180
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 128 TTCTGACCTGTTTACAGCAGAGTGTGATGATGTTTCAAGATGATGGGATGGAGAA 187
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 181 AATACCTCTGACCTCTCCAGGCGAGACAAAGAAAGCGCAAGATGCTGACCAACTT 240
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 188 AACACCTCTGACCTCTCCAGGCGAGACAAAGAAAGCGCAAGATGCTGACCAACTT 247
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 241 GGAACCCAGCCCAAGAAAGAAACGTAATGCTGAGAGAAATTAATATATA 300
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 248 GGAACCCAGCCCAAGAAAGAAAGAAACGTAATGCTGAGAGAAATTAATATATA 307
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 301 GAAGAACTTGCAGAGTGTATTTTGCATTTTATGATATAGCAACTTTACTTCAA 360
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 308 GAGAGAGTGGCCGAGTGTATCTTGCATTTATGATATAGCAACTTTACTTCAA 367
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 361 CCTGACCAATGTGCAATCTTAAAGAAAGCTGACCAATTCCTGACATCAAGAACAA 420
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 368 CCTGACCAATGTGCAATCTTAAAGAAAGCTGACCAATTCCTGACATCAAGAACAA 427
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 421 GAGAAACACAGAGTGCATGATGATGATGATGATGATGATGATGATGATGATGATG 480
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 428 GAGAAACACAGAGTGCATGATGATGATGATGATGATGATGATGATGATGATGATG 487
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 481 CAGGCTGTCATCGACAGAGTGGCGCTATGATGCTTGAAGCCCTTGATGGTTC 540
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 488 CAGGCTGTCATCGACAGAGTGGCGCTATGATGCTTGAAGCCCTTGATGGTTC 547
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 541 TTCTTTTATGTAACCTGGAAGGCAAGCTGTGTTTGTGACAGAAATGTGACAGATAT 600
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 548 TTCTTTTATGTAACCTGGAAGGCAAGCTGTGTTTGTGACAGAAATGTGACAGATAT 607
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 601 CTAAGGTATACCAAGAGAGCTG 624
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 608 CTAAGGTATACCAAGAGAGCTG 631
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

RESULT 14  
LOCUS BG987973 566 bp mRNA linear EST 13-JUN-2001  
DEFINITION MR2-HT1162-110101-001-b05 HT1162 Homo sapiens cDNA, mRNA sequence.  
ACCESSION BG987973  
VERSION BG987973.1 GI:14392043  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 566)  
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,  
Nagel,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,P.F.,  
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,  
Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,  
M.D., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and  
Simpson,A.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
MEDLINE 20202663  
COMMENT Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=MR2-HT1162-  
110101-001-b05&st3=2001-01-11&st4=1)  
Seq primer: puc 18 forward  
High quality sequence stop: 281.  
Location/Qualifiers

FEATURES  
source  
1..566  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="HT1162"  
/dev\_stage="Adult"  
/note="Organ: head,neck; Vector: puc18; Site.1: SmaI;  
Site.2: SmaI; A mini-library was made by cloning products  
derived from ORESSES PCR (U.S. Letters Patent application  
No.196,716 - Ludwig Institute for Cancer Research)  
profiles into the puc 18 vector. Reverse transcription of  
tissue mRNA and cDNA amplification were performed under  
low stringency conditions."

BASE COUNT 166 a 147 c 174 g 79 t  
ORIGIN

Query Match 8.0%; Score 492.4; DB 10; Length 566;  
Best Local Similarity 94.2%; Pred. No. 3.2e-105;  
Matches 533; Conservative 0; Mismatches 31; Indels 2; Gaps 2;

```
QY 1805 TAGGGTATGCTGGGCTTACACAGACTTAAATGGCAATTTGCAAACTCCCACTTA 1864
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1 TAGGGTATGCTGGGCTTACACAGACTTAAATGGCAATTTGCAAACTCCCACTTA 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1865 ATATGATTCCTCCCGCCTACAGAGATGGAGGCTGAGCTCAAAAGACTGTTTGAC 1924
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 61 ATATGATTCCTCCCGCCTACAGAGAGGAGAGAGTGGAGCTCAAAAGACTGTTTGAC 120
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1925 TATATGGGAGCCCTCTGAGAGTACAACCTGAGACAAGAGAGAGAGCTGCATCTGGAG 1984
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 121 TATATGGGAGCCCTCTGAGAGTACAACCTGAGACAAGAGAGAGAGCTGCATCTGGAG 180
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1985 AGCAAAAGAAACAA-TGACCCCAACCTGCCCGGCGCTGAGAGAGAGCTGAC 2043
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 181 AGCAAAAGAAACAAAGTGAACCCCAACCGCCCGGCGCTGAGAGAGAGCTGAC 240
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2044 GGGCAGAGCAGAGCTGCA-TGACAGCAAGAGGCGAGACCAACTCCTGAGCTGCGACAC 2102
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 241 GGGCAGAGCAGAGGCGAGTACACAGCAAGGGCGAGACCAACTGCGAGCGGAGCAC 300
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2103 CAATCTGATCAGATGAGAGCCCTCGCCCTTACAGCAGCTTTTGTGCGATACAAACAAGA 2162
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 301 CAATCTGATCAGATGAGAGCCCTCGCCCTTACAGCAGCTTTTGTGCGATACAAACAAGA 360
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2163 CTCACAGAGTACTGCTGCTGCTGTGCTGTACATGAGAACTGCTGCTCAAGAGAGCA 2222
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 361 CTCACAGAGTACTGCTGCTGCTGTGCTGTACATGAGAACTGCTGCTCAAGAGAGCA 420
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2223 TAAATTTTGCACAGACTCTTGCAGAGCAGACAGTCCCTGCTGAGACTGGCCAAAGTTAAC 2282
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 421 TAAATTTTGCACAGACTCTTGCAGAGCAGACAGTCCCTGCTGAGACTGGCCAAAGTTAAC 480
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2283 AGCAGAAAGCAGAGCAGAAAGCTGAGCGAGAGTCCAGAGCAGACAGCTCTGATAGAGA 2342
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 481 AGCAGAAAGCAGAGCAGAAAGCTGAGCGAGAGTCCAGAGCAGACAGCTCTGATAGAGA 540
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2343 AGTACTATTAAACAGAGCCGCTGA 2368
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 541 AGTACTATTAAAGAGAGCGCGGGA 566
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	TITLE	JOURNAL	COMMENT	FEATURES	BASE COUNT	ORIGIN
BE031423	BE031423	129986	BE031423	BE031423.1	GI:8326432	EST	Design and use of two pooled tissue normalized cDNA libraries for EST discovery in swine	Unpublished (2000)	Contact: Smith TPL USDA, ARS, US Meat Animal Research Center PO Box 166, Clay Center, NE 68933-0166, USA Tel: 402 762 4366 Fax: 402 762 4390 Email: smithemall.marc.usda.gov	Single pass sequencing. Bases called and alt_trimmed with phred v0.980304.e. Vector identified by cross_match with the -minscore 18 and -mismatch 12 options.	170 a 110 c 140 g 125 t	
BE031423	129986	MARC	1Pig Sus scrofa cDNA 5', mRNA sequence.	1 (bases 1 to 545)	Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.							
BE031423	129986	MARC	1Pig Sus scrofa cDNA 5', mRNA sequence.	1 (bases 1 to 545)	Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.							
BE031423	129986	MARC	1Pig Sus scrofa cDNA 5', mRNA sequence.	1 (bases 1 to 545)	Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.							
BE031423	129986	MARC	1Pig Sus scrofa cDNA 5', mRNA sequence.	1 (bases 1 to 545)	Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.							
BE031423	129986	MARC	1Pig Sus scrofa cDNA 5', mRNA sequence.	1 (bases 1 to 545)	Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.							
BE031423	129986	MARC	1Pig Sus scrofa cDNA 5', mRNA sequence.	1 (bases 1 to 545)	Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.							
BE031423	129986	MARC	1Pig Sus scrofa cDNA 5', mRNA sequence.	1 (bases 1 to 545)	Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.							
BE031423	129986	MARC	1Pig Sus scrofa cDNA 5', mRNA sequence.	1 (bases 1 to 545)	Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.							
BE031423	129986	MARC	1Pig Sus scrofa cDNA 5', mRNA sequence.	1 (bases 1 to 545)	Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.							
BE031423	129986	MARC	1Pig Sus scrofa cDNA 5', mRNA sequence.	1 (bases 1 to 545)	Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.							
BE031423	129986	MARC	1Pig Sus scrofa cDNA 5', mRNA sequence.	1 (bases 1 to 545)	Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.							
BE031423	129986	MARC	1Pig Sus scrofa cDNA 5', mRNA sequence.	1 (bases 1 to 545)	Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.							
BE031423	129986	MARC	1Pig Sus scrofa cDNA 5', mRNA sequence.	1 (bases 1 to 545)	Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.							
BE031423	129986	MARC	1Pig Sus scrofa cDNA 5', mRNA sequence.	1 (bases 1 to 545)	Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.							
BE031423	129986	MARC	1Pig Sus scrofa cDNA 5', mRNA sequence.	1 (bases 1 to 545)	Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.							
BE031423	129986	MARC	1Pig Sus scrofa cDNA 5', mRNA sequence.	1 (bases 1 to 545)	Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.							
BE031423	129986	MARC	1Pig Sus scrofa cDNA 5', mRNA sequence.	1 (bases 1 to 545)	Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.							
BE031423	129986	MARC	1Pig Sus scrofa cDNA 5', mRNA sequence.	1 (bases 1 to 545)	Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.							
BE031423	129986	MARC	1Pig Sus scrofa cDNA 5', mRNA sequence.	1 (bases 1 to 545)	Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.							
BE031423	129986	MARC	1Pig Sus scrofa cDNA 5', mRNA sequence.	1 (bases 1 to 545)	Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.							
BE031423	129986	MARC	1Pig Sus scrofa cDNA 5', mRNA sequence.	1 (bases 1 to 545)	Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.							
BE031423	129986	MARC	1Pig Sus scrofa cDNA 5', mRNA sequence.	1 (bases 1 to 545)	Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.							
BE031423	129986	MARC	1Pig Sus scrofa cDNA 5', mRNA sequence.	1 (bases 1 to 545)	Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.							
BE031423	129986	MARC	1Pig Sus scrofa cDNA 5', mRNA sequence.	1 (bases 1 to 545)	Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.							
BE031423	129986	MARC	1Pig Sus scrofa cDNA 5', mRNA sequence.	1 (bases 1 to 545)	Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.							
BE031423	129986	MARC	1Pig Sus scrofa cDNA 5', mRNA sequence.	1 (bases 1 to 545)	Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.							
BE031423	129986	MARC	1Pig Sus scrofa cDNA 5', mRNA sequence.	1 (bases 1 to 545)	Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.							

QY	610	AACCAAGAAAGCCTGATACACAAAGTGTATATATGACATCTTGACATGTTGGGAGCCACACG	669
Db	362	AACCGAGAGACCTGATATACAAAGCGCTATATATACATCTTGACATGTTGGGAGCCACACG	421
QY	670	GAATTTGTCAAAAACCTCTGCGCAAGTCTATAGTAAATGGGGGATCTTGGCTGCGGAA	729
Db	422	GAATTTGTCAAAAACCTCTGCGCAAGTCTATAGTAAATGGGGGATCTTGGCTGCGGAA	481
QY	730	CCTCCGAGCGGAGACGCCATACCTTCAATTTGTGGATGCTGATGAACCTTTACCTGAT	789
Db	482	CCTCCGAGCGGAGACGCCACACTTTCATATGTGGATGTTGGTGAACCTCTGCTGAT	541
QY	790	TCAG 793	
Db	542	CCAG 545	
RESULT	16		
LOCUS	BE031427	543 bp	mRNA linear EST 09-JUL-2000
DEFINITION	129992 MARC 1P1G Sus scrofa cDNA 5', mRNA sequence.		
ACCESSION	BE031427		
VERSION	BE031427.1	GI:8326436	
KEYWORDS	EST.		
SOURCE	plg.		
ORGANISM	Sus scrofa		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus. 1 (bases 1 to 543) Fahrtenkrug,S.C., Fieking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W., and Keeler,J.W. Design and use of two pooled tissue normalized cDNA libraries for EST discovery in swine Unpublished (2000) Contact: Smith TPL USDA, ARS, US Meat Animal Research Center PO Box 166, Clay Center, NE 68933-0166, USA Tel: 402 762 4366 Fax: 402 762 4390 Email: smith@email.marc.usda.gov Single pass sequencing. Bases called and alt-trimmed with phred v0.980904.e. Vector identified by cross_match with the -mnscore 18 and -minmatch 12 options. PCR Primers FORWARD: AGGAACAGCATATGACCAT BACKWARD: GTTTTCCAGCTACGACG Plate: 66 row: L column: 9 Seq primer: ATTTAGGTGACCATATAG.		
FEATURES	source	1..543	Location/Qualifiers
		/organism="Sus scrofa"	
		/db_xref="taxon:9823"	
		/clone_lib="MARC 1P1G"	
		/tissue_type="pooled"	
		/lab_host="DH10B"	
		/note="Vector: pCMV SPORT6; Site.1: XbaI; Site.2: XhoI; Library made from pooled tissue from day 11, 13, 15, 20, and 30 embryos."	
BASE COUNT	168 a	110 c	140 g 125 t
ORIGIN			
Query Match.	7.9%	Score 487.6;	DB 9; Length 543;
Best Local Similarity	93.7%;	Pred. No. 4.4e-104;	
Matches 508:	Conservative 0;	Mismatches 34;	Indels 0; Gaps 0;
QY	250	CCCAAAAGGAACATCGAAAACGTATGCTGTAACGAAAATTAATATATAGAGAACTT	309
Db	2	CCCAAGAGAGACACATGAGAAACGATATGCTGACAGGAAAAATTAATACATGGAAGAACTT	61
QY	310	GCAGAGTGTGATTTTGACAATTTTAATGATATAGACAACTTTAACTTCAAACTGACAA	369

Db	62	GCAGAGCGATTTTTTGGCAATTTTAACGATATAGACAACTTTAACTTCAAACTGACAAA	121
Oy	370	TGTGCAATCTTTAAAGAAAATGTGTGAAGCAAAATTCGTACAGATCAAAAGACAGAAAAGCA	429
Db	122	TGTGCAATCTTTAAAGAAAATGTGTGAAGCAAAATTCGTACAGATCAAAAGACAGAAAAGCA	181
Oy	430	GCAGTGTGCAACATATGATGAAAGTGCAGAAAGTACAGTATCCTCTACAGAGGACAGGTGTC	489
Db	182	GCAGTGTGCAACATATGATGAAAGTGCAGAAAGTACAGTATCCTCTACAGAGGACAGGTGTC	241
Oy	490	ATCGACAAGAGATGCGCTGGGCGCTATGATAGCTTGAGGCGCCCTGATGGGTCTTCTTTGTA	549
Db	242	ATCGACAAGAGACGCGCTGGGCGCTATGATAGCTTGAGGCGCCCTGATGGGTCTTCTTTGTA	301
Oy	550	GTTGAACCTGTGAAGGCGAACGTTGTGTTTGTGTGCAGAGAAATGTACACAGATATCTAAGTAT	609
Db	302	GTTGAACCTGTGAAGGCGAACGTTGTGTTTGTGTGCAGAGAAATGTACACAGATATCTAAGTAT	361
Oy	610	AACCAAGAAAGAGCTATATGACAAAGTATATATACATCTTCCATGTTGGGACACACACG	669
Db	362	AACCAAGAAAGAGCTATATGACAAAGTATATATACATCTTCCATGTTGGGACACACACG	421
Oy	670	GAATTTGTCAAAAAAAGCTGTGCGCAAGTCTATAGTAATGGGGGAGATCTTGTCGGCGAA	729
Db	422	GAATTTGTCAAAAAAAGCTGTGCGCAAGTCTATAGTGAAGCGGGGATCTTGTCGTGTGTGA	481
Oy	730	CCTCCGAGCGGAACAGCCATACCTTCAATGTGTGGATGTGGTAAAACTTTACCTGAT	789
Db	482	CCTCCGAGCGGAGAGACGACACACTTTCAAATGTGTGGATGTGGTGAACACTCTGCTGTAT	541
Oy	790	TC 791	
Db	542	CC 543	

RESULT	17
LOCUS	BB628347
DEFINITION	BB628347 RIKEN cDNA full-length enriched, adult male urinary bladder muscle.
ACCESSION	BB628347
VERSION	BB628347.1
KEYWORDS	GI:16465803
SOURCE	EST.
ORGANISM	house mouse. Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 655)
AUTHORS	Arakawa,T., Carlini,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hiramoto,K., Horii,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Koude, .M., Koyama,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasakih, .D., Shibata,K., Shingawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.
TITLE	RIKEN Mouse ESTs (Arakawa,T. et al. 2001)
JOURNAL	Unpublished (2001)
COMMENT	Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel.: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@gsc.riken.go.jp,  
URL: <http://genome.gsc.riken.go.jp/>  
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoch, K.,  
M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new  
genes. *Genome Res.* 10 (10), 1617-1630 (2000)  
wagii, K., Fujiwara, S., Inoue, K., Togawa, Y., Itawa, M., Ohara, E.,

Matsubara, M., Konekawa, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunaga, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.

Riken integrated sequence analysis (RISA) system -384-format sequencing pipeline with 384 multichannel sequencer. *Genome Res.* 10 (11), 1757-1771 (2000)

Komuro, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara, Y. and Hayashizaki, Y.

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. *Genome Res.* 11 (2), 281-289 (2001)

Kondo, S., Shingawa, A., Saito, T., Kiyosawa, H., Yamashita, I., Aizawa, K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K. and Hayashizaki, Y.

Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences. *Mamm. Genome*. 12, 673-677 (2001)

Please visit our web site (<http://genome.ysc.riken.go.jp>) for further details.

e mouse tissues.

FEATURES

SOURCE

```

/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="9530095N19"
/clone_id="RIKEN full-length enriched, adult male urinary bladder"
/sex="male"
/tissue_type="urinary bladder"
/dev_stage="adult"
/lab_host="DH10B"
/notes="Site_1: SalI; site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5', GAGAGAGAGAGATCCAGAGCTCTTTTCTTTTCTTTT 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 20.0 and subtraction to Rot = 370.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAAATTAAATTCACCCCCCCCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified Bluescript KS(+) after bulk excision from Lambda FIC I."

```

[illegible]

Db	241	CTGAGAGCATCCTGGCTTTAAGACGAGGAGATGTTTGGCAGCCAGAAACAGGACGCTTCG	300
Oy	3332	GCAGTTTCTCCAGATGACTTGTCTATGTCCACATCTCTGCAGCTGATGTCCGAGTATGAG	3391
Db	301	GCAGCTCCCTCGATGACCTGCTGTCTCCACATCTCTGCAGCAGAGTCTGCCAAGCGATGAGG	360
Oy	3392	GAGCTCTCCTGGACAGCTATCTGGCTTCGGGAATTTGATGATGCCGAGGAGATTTG	3451
Db	361	GCGCTCTTCTGACCAAGCTGATCTGGGCTTCGGAACTTCGATGCGCTTGGAGGATTTG	420
Oy	3452	ATAGAGCCTTAAAGATPACCCGAACTGTCTCAGCCAGCAAGCCATAGATCAGAAACAT	3511
Db	421	ATAGAGCTCTGGGGATACCAGAACTGTGTCTCAGCCAGCAAGCCATGTGATGCAAGACAT	480
Oy	3512	TCTCAAGTCAGGATTTCCAAATCATGATCTGTGGAGCAGAAAGGCGCCGTTTCCACAGCACT	3571
Db	481	TCTCAAGTCAGAGATCCACCATPAAATGCTGTGGACAAAGGCCGCCCTTTTCCACAGCACT	540
Oy	3572	ATGCATCTTCAGGCACAAATGGCCCGAGGGTAGTATTCTCCATTCGAAG-ATCCAAACTTT	3630
Db	541	ACCCCTTTTAAACACAAATGGGCGAGGGGTGCTATTAATCCCTCGCAAGATCCAAACTTT	600
Oy	3631	CACACCATGGGACAGGGGCTA 3652	
Db	601	TACACCATGGGAAAGCGGGCCA 622	

RESULT	18		
BG835556			
LOCUS	BG835556	559 bp	mRNA linear
DEFINITION	354744 MARC IPiG Sus scrofa cDNA 5', mRNA sequence.		EST 25-MAY-2001

KEYWORDS	EST.
SOURCE	plg.
ORGANISM	
	<i>Sus scrofa</i>
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
	Mammalia; Eutheria; Cetartiodactyla; Suidae; Suidae; Sus.
REFERENCE	1 (bases 1 to 559)
AUTHORS	Falhekreiyu,S.C., Frcking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,

TITLE	Design and use of two pooled tissue normalized cDNA libraries for EST discovery in swine
JOURNAL	Unpublished (2000)
COMMENT	Contact: Smith TPL

PO Box 166, Clay Center, NE 68933-0166, USA  
Tel: 402 762 4366  
Fax: 402 762 4390  
Email: [smith@emall.marc.usda.gov](mailto:smith@emall.marc.usda.gov)  
Single pass sequencing. Bases called and alt-trimmed with phred  
v0.980904.e. Vector identified by cross-match with the -mismatch 16  
and -minmatch 12 options.

```

FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTTCCCACTACACGACG
Plate: 117 row: D column: 21
Seq primer: ATTTAGTGACACTATAG.
FEATURES
    source          1..559

```

BASE COUNT	171 a	139 c	149 g	100 t
ORIGIN				

Query Match	7.8%	Score 483;	DB 10;	Length 559;
Best Local Similarity	92.7%	Pred. No. 5.4e-103;		
Matches 518; Conservative	0;	Mismatches 40;	Indels 1;	Gaps 1;

2170 GGTAGCTTGCTGTTCTGGGTCTACACATGGAACCTGCTCAA - GGAGAGCATAAAT 2228

Db 1 GGCAGCCTGCTGGTCTGGGGGACACACGGAACTCGCTCAAGGGAAAAGCATAGAT 60

QY 2229 TTTCACACAGCTCTTGACAGACACAGTTCCTCTGTGACTTGCCCAAGTTAACAGCAGA 2288

Db 61 TTTCACACACTGTGGCAGGATAGCAGCTCTCCTGTTGGACTTGGCGAAGTTAACTCGCGA 120

[illegible][illegible]

181 GATTAAAGCAGAGCCAGTGAGCCCAAGAGAAAGACAAATGCACTACTACCGGCTATTGGCTT 240

2409 AGATAAGATGATACTAAAGATATTGGTTTACCAGAAATAACCCCAACTTGAGAGACT 2468

Db 241 CGATAAGACGACTAAGATATCGGTTACCGGAATTAACCCCAACTTGAGCGACT 300

2469 GGACAGTAAGACAGATCCTGCAGTAACACAAATTAATAGCAATGAAAACTGAGAAGCA 2528

Db 301 GGACAGTAAGACGGACCCCTGCCAGTACACAAATTATAGCTATGAAAACGTGAGAAGGA 360

QY 2529 GGAGATGAGCCTTTGAGCCTGTGTGACCAAGCCTGGCAGTGAAGCTGGACAACCTTGGAGGAGAT 2588

Db 361 GGAGATGAGCITTTGAGCCTAGTGACACGCGCTGGCAGTGAAGCTGGACAACITGGAGGAGAT 420

QY 2589 TTTGGATGATTTCACAGATAGTCAATTACCCACAGCTTTCCACAGACACGAGSCCAGGGCC 2618

DB 421 TTGGATGATTTGGAGAATAGTCAATTAACACAGCCTTTCCAGACACGAGGCCAAGCC 400

Year	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025	2026	2027	2028	2029	2030	2031	2032	2033	2034	2035	2036	2037	2038	2039	2040	2041	2042	2043	2044	2045	2046	2047	2048	2049	2050	2051	2052	2053	2054	2055	2056	2057	2058	2059	2060	2061	2062	2063	2064	2065	2066	2067	2068	2069	2070	2071	2072	2073	2074	2075	2076	2077	2078	2079	2080	2081	2082	2083	2084	2085	2086	2087	2088	2089	2090	2091	2092	2093	2094	2095	2096	2097	2098	2099	2100
1990	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100											

[illegible]

541 AAACAGCCCCGTTACACCTT 559

RESULT 19  
AW465013/C

LOCUS	AM403013	342 bp	linear	ESI 24 FEB 2000
DEFINITION	BP230017B10B11 Soares normalized bovine placenta Bos taurus cdna			

ACCESSION AM465013  
 UNDPSTN 1 CT. 7035101

KEYWORDS	EST.
SCIENCE	COW

ORGANISM    Bos taurus  
Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:

Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.

REFERENCE	AUTHORS
1 (bases 1 to 342)	Lewin, H.A., Soares, M.B., Rebeiz, M., Pardini, J., Liu, L. and Larson

TITLE	U.S.N.
Bovine ESTS	

COMMENT  
Contact: Lewin, H. A.  
at Wash. Center for Communitarian and Functional Economics

University of Illinois at Urbana-Champaign  
340 Edward R. Madigan Laboratory 1201 W. Gregory Dr. Urbana, IL

61801, USA  
Tel: 217 333 5998

Fax: 217 244 5617  
Email: h-lewin@uiuc.edu

Funding for cattle ESI sequencing was provided by the USDA National Research Initiative, Animal Genome Resource Grant AG 99-3205-8534



Matches 480; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2126 CGCCCTTACGACCTCTTTGCGGATACAAACAAAGACTCCAGAGTAGCTTCGCTT 2185  
 |||||||  
 Db 5 CGCCCTTACGACCTCTTTGCGGATACAAACAAAGACTCCAGAGTAGCTTCGCTT 64  
 |||||||

QY 2186 CTGGGCTCTACACTGGAACCTCGCTCAAGAGAAAGCATTAATTTTGCACAGACTCTTGC 2245  
 |||||||  
 Db 65 CTGGGCTCTACACTGGAACCTCGCTCAAGAGAAAGCATTAATTTTGCACAGACTCTTGC 124  
 |||||||

QY 2246 AGGACAGACAGTTCCCTGTGGACTTGCCCAAGTTAACAGAGAGACACAGGCAAGAC 2305  
 |||||||  
 Db 125 AGGACAGACAGTTCCCTGTGGACTTGCCCAAGTTAACAGAGAGACACAGGCAAGAC 184  
 |||||||

QY 2306 TGAAGCAGAGATCCAGACAGACAGCTCTGATCAGAAAGTACTATTAAACAAGAGCCGG 2365  
 |||||||  
 Db 185 TGAAGCAGAGATCCAGACAGACAGCTCTGATCAGAAAGTACTATTAAACAAGAGCCGG 244  
 |||||||

QY 2366 TGAAGCAGAGAGAAAGAGAAATGCACTACTTGGCTATTGCTAGATAAAGATGATCTA 2425  
 |||||||  
 Db 245 TGAAGCAGAGAGAAAGAGAAATGCACTACTTGGCTATTGCTAGATAAAGATGATCTA 304  
 |||||||

QY 2426 AAGATATGTTGTTACCAAGAAATTAACCCCAACTTGGAGAGCTGGACAGTAAGACAGATC 2485  
 |||||||  
 Db 305 AAGATATGTTGTTACCAAGAAATTAACCCCAACTTGGAGAGCTGGACAGTAAGACAGATC 364  
 |||||||

QY 2486 CTGCCAGTAAACAAATTAATAGCAATGAAACTGAGAGAGAGAGATAGCTTTGAGC 2545  
 |||||||  
 Db 365 CTGCCAGTAAACAAATTAATAGCAATGAAACTGAGAGAGAGAGATAGCTTTGAGC 424  
 |||||||

QY 2546 CTGGTGACACAGCTGTGAGCTGAGACACATTGGAGAGATTTGGATGATTTGCACA 2605  
 |||||||  
 Db 425 CTGGTGACACAGCTGTGAGCTGAGACACATTGGAGAGATTTGGATGATTTGCACA 484  
 |||||||

RESULT 21  
 BM090425 555 bp mRNA linear EST 19-NOV-2001  
 LOCUS 505844 MARC 3BOV Bos taurus cDNA 5', mRNA sequence.  
 ACCESSION BM090425  
 VERSION BM090425.1 GI:17001053  
 KEYWORDS EST.  
 SOURCE cow.  
 ORGANISM Bos taurus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 Bovidae; Bovinae; Bos.  
 1 (bases 1 to 555)  
 Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,  
 Casas,E., Wray,J.B., White,J., Cho,J., Fahrenkrug,S.C., Bennett,  
 G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,  
 Perlea,G., Holt,I., Karameycheva,S., Liang,F., Quackenbush,J. and  
 Keefe,J.W.  
 Sequence evaluation of four pooled-tissue normalized bovine cDNA  
 libraries and construction of a gene index for cattle  
 Genome Res. 11 (4), 626-630 (2001)  
 21180013  
 Contact: Smith TPL  
 USDA, ARS, US Meat Animal Research Center  
 PO Box 166, Clay Center, NE 68933-0166, USA  
 Tel: 402 762 4366  
 Fax: 402 762 4390  
 Email: smith@email.marc.usda.gov  
 Single pass sequencing. Bases called and alt-trimmed with phred  
 v0.980904.e. Vector identified by cross\_match with the -mnscore 18  
 and -mismatch 12 options.  
 PCR Primers  
 FORWARD: AGGAACACAGCTATGACCAT  
 BACKWARD: GTTTCACAGTCACGACG  
 Plate: 98 row: F column: 6  
 Seq primer: ATTTAGTACACTATAG.  
 Location/Qualifiers  
 1 .555

Query Match 7.8%; Score 478.2; DB 10; Length 555;  
 Best Local Similarity 91.4%; Pred. No. 7.3e-102;  
 Matches 507; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

QY 2787 AGGGCAACGTGGGACAGTTATTTGCCAACAGCAATTTACCACTTGACATACATTGCAAG 2846  
 |||||||  
 Db 1 AGGGCAACGTGGGACAGTTATTTGCCAACAGCAATTTCCACTGACATACATTGCAAG 60  
 |||||||

QY 2847 CCCACTGTGCTGACCTTTCCACCAATCAGAAACAGTAGTCCCTACTCATGATATCC 2906  
 |||||||  
 Db 61 CCCAAGGGGCGCTGACCTTTCCACCAATCAGAAACAGTAGTCCCTACTCATGATATCC 120  
 |||||||

QY 2907 TCAGCCAGGAATGATGGTAAATCAAGGATGATAGGAAACCAAGAAATTTAGGAACAG 2966  
 |||||||  
 Db 121 TCAGCCAGGAATGATGGTAAATCAAGGATGATAGGAAACCAAGAAATTTAGGAACAG 180  
 |||||||

QY 2967 TAGCAGAGATGATGTTGTAACAGTCTTCTGAGCTTACTATGCTCTGGAAGATGGGC 3026  
 |||||||  
 Db 181 TAGCAGAGATGATGTTGTAACAGTCTTCTGAGCTTACTATGCTCTGGAAGATGGGC 240  
 |||||||

QY 3027 ACCGACAGTGTGGCTGTGAGAGTCACTGTGTGTCTACACAGTGCATGAACCGGCG 3086  
 |||||||  
 Db 241 GCCCAGAGTCTGCGGTGAGGTCACTGTGTGTCTACACAGTGCATGAACCGGCG 300  
 |||||||

QY 3087 AGTCCAAAGAGATGATTTGGGAACCCAGCAGCAGATCCCATGAGGCCAGACCA 3146  
 |||||||  
 Db 301 AGTCCAAAGAGATGATTTGGGAACCCAGCAGCAGATCCCATGAGGCCAGACCA 360  
 |||||||

QY 3147 GCTGGCCAAAGACACAGCCTTCAGTCAAGTCAATGATTAAGGCGCATGGAATTA 3206  
 |||||||  
 Db 361 GCTGGCCAAAGACACAGCCTTCAGTCAAGTCAATGATTAAGGCGCATGGAATTA 420  
 |||||||

QY 3207 GATGACATAGGGGGGAGCTCAGTATAGCCAAACAAGCTCTCCAAATCAGACTGCC 3266  
 |||||||  
 Db 421 GATGACATAGGGGGGAGCTCAGTATAGCCAAACAAGCTCTCCAAATCAGACTGCC 480  
 |||||||

QY 3267 ATGGCTGAAAGCATCTCTGCTATAGACACAGGCGTCTTTGCCAGCCAAACAGCAGCC 3326  
 |||||||  
 Db 481 ATGGCTGAAAGCATCTCTGCTATAGACACAGGCGTCTTTGCCAGCCAAACAGCAGCC 540  
 |||||||

QY 3327 ATTTGGCAGTTCYC 3341  
 |||||||  
 Db 541 ATTTGGCAGTTCYC 555  
 |||||||

RESULT 22  
 BG172538 777 bp mRNA linear EST 06-FEB-2001  
 LOCUS BG172538  
 DEFINITION 60233736F1 NCI\_CGAP\_Mam1 Mus musculus cDNA clone IMAGE:4456760 5',  
 mRNA sequence.  
 ACCESSION BG172538  
 VERSION BG172538.1 GI:12679241  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 777)  
 NIH-MGC http://mgc.nci.nih.gov/.  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 TITLE





```
Db 4 GGAGAGCGGGTCTCATTTAGGTCATGCTGTGGCTTCAACACACCTAAATAAGGGCAATTTC 63
OY 1849 CAAATCTCCCATTTATATGATATCCCTCCCACTCAGCAA-GATGGAGAGCTTGACATC 1907
Db 64 CAAACTCCCATTTATATGATATCTCCCACTCAGCAAAGGAGGGAGGCTGGACTC 123
OY 1908 AAAAGACTGTTTGGACTATATGGGAGCCCTCTGAAGTACAACTGGACACAGAGAG 1967
Db 124 AAAAGAGCGGTTGGACTATATGGGAGCCCTCTGAAGGTACAACTGGACACAGAGAG 183
OY 1968 CAGCTGCCATCTTGAGAGAGCAAAAGAACAA-TGACCCCAACTTGCCCGGCGCTGA 2026
Db 184 CAGCTGCCATCTTGAGAGAGCAAAAGAACAAAGTGAACCCCAACTTGCCCGGCGCTGA 243
OY 2027 GCAGTGA--GAGAGCTGACGGGAGAGACAGACTGATGACAGCAAGGGGAGACCAACT 2084
Db 244 GCAGTACAGAGAGCGTGACGGGAGAGACAGACTGATGACAGCAAGGGGAGACCAACT 303
OY 2085 CCTGACAGCTGCTGACCAACCAATCTGATCAGA-TGGAGCCCTGCGCCTTGAGCCAGCTCTT 2143
Db 304 CCGGCGGCTGCTGATACCAACCAATCTGATCAGAGTGAAGCCCTGCGCCTTGAGCCAGCTCTT 363
OY 2144 TGTGGATFACAAACAAAGACTCCACAGGTAGCTTGCCTGCTTGGGTCTACACATGGAA 2203
Db 364 GGTGGATFACAAACAAAGACTCCACAGGTAGCTGCGCGGCTTGGGTCTACACATGGAA 423
OY 2204 CCTGCGTCAAGGAGAGCATAAATTTTGGACAGACTCTTGGACAGAGAGAGTTCCTCTG 2263
Db 424 CCTGCGTCAAGGAGAGCATAAATTTTGGACAGACTCTTGGACAGAGAGAGTTCCTCTG 483
OY 2264 TGGACTTGGCCAAAGTTAACAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGTCCAGCA 2323
Db 484 TGGACTTGGCCAAAGTTAACAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGTCCAGCA 543
OY 2324 GCACACTCTCTGATCA-GAAGTGAATTTAAACAAGAGCGCGGTGAGCC 2371
Db 544 GCCAGGCTCTGATCAGAAAGTGAATTTAAACAAGAGAGAGCGGTGAGAGC 592

RESULT 24
AA744255 467 bp mRNA linear EST 23-JAN-1998
DEFINITION ny62911.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1282916 3',
ACCESSION AA744255
VERSION AA744255.1 GI:2785005
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
TITLE 1 (bases 1 to 467)
JOURNAL NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
COMMENT National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
Ph.D., Gerald Marti, M.D.
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/dbp/image/image.html
Insert length: 1583 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 391.
Location/Qualifiers
1..467
```

```
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1282916"
/clone.lib="NCI CGAP GCB1"
/tissue_type="germinal center B cell"
/lab_host="DH10B"
/note="Vector: pTR3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from human tonsillar cells enriched for
germinal center B cells by flow sorting (CD20+, IgD-),
provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
(NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was
primed with a Not I - 0190(dt) primer
15'-TGTTACCAATGGAAGTGGAGGAGGCGGCTCATTTTGTGTTTGTGTTT-3'
]. Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pTR3 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo."
```

BASE COUNT 151 a 106 c 114 g 96 t

ORIGIN

Query Match 7.5%; Score 462.2; DB 9; Length 467;  
Best Local Similarity 99.4%; Pred. No. 4.2e-98;  
Matches 464; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
OY 2229 TTGGCAGAGACTCTGACAGAGAGAGTTCCTCGTGGACTTGCCCAAGTTAACAGACAGA 2288
Db 1 TTGTGACAGACTCTTGACAGAGAGAGTTCCTCGTGGACTTGCCCAAGTTAACAGACAGA 60
OY 2289 AGCCACAGGCAAGACCTGAGAGAGAGAGTCCAGACAGAGTCTGATCAGAAAGTAC 2348
Db 61 AGCCACAGGCAAGACCTGAGAGAGAGAGTCCAGTACAGAGTCTGATCAGAAAGTAC 120
OY 2349 TATTAAACAAGAGCGCGGTGAGCCCAAGAAAGAGAGATGCACTACTTGGCTATTGGCT 2408
Db 121 TATTAAACAAGAGCGCGGTGAGCCCAAGAAAGAGAGATGCACTACTTGGCTATTGGCT 180
OY 2409 AGATTAAGATGATACATAAAGATATGTTTACAGAGAAATACCCCAAACTTGAAGACT 2468
Db 181 AGATTAAGATGATACATAAAGATATGTTTACAGAGAAATACCCCAAACTTGAAGACT 240
OY 2469 GGACAGTAAAGAGATCTTCCAGCTAACAACAAATTAATGCAATGAAACTGAGAAAGA 2528
Db 241 GGACAGTAAAGAGATCTTCCAGCTAACAACAAATTAATGCAATGAAACTGAGAAAGA 300
OY 2529 GGAGATGAGCTTTGAGCTGTGACACAGCCTGGCAGTGAAGCTTGACAACTTGGAGAGAT 2588
Db 301 GGAGATGAGCTTTGAGCTGTGACACAGCCTGGCAGTGAAGCTTGACAACTTGGAGAGAT 360
OY 2589 TTGGATGATTTTGCAGATATGTCATTAATTCACAGAGCTTTTCCAGACAGAGGCGCGC 2648
Db 361 TTGGATGATTTTGCAGATATGTCATTAATTCACAGAGCTTTTCCAGACAGAGGCGCGC 420
OY 2649 CCCTGCTGATCAGTTGACAAGCAAGCCATCATGATGACCTCATGC 2695
Db 421 CCCTGCTGATCAGTTGACAAGCAAGCCATCATGATGACCTCATGC 467

RESULT 25
AM504723 466 bp mRNA linear EST 02-MAR-2000
LOCUS AM504723
DEFINITION UI-HE-BNO-alm-d-01-0-UI.r1 NIH_MGC_50 Homo sapiens cDNA clone
ACCESSION AM504723
VERSION AM504723.1 GI:7142390
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 466)
```



AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
 Eco RI site shown at the beginning of the sequence.  
 Tissue Procurement: Louis M. Staudt, M.D., Ph.D.  
 cDNA Library Preparation: M.B. Soares Lab  
 cDNA Library Arrayed by: M.B. Soares Lab  
 DNA Sequencing by: M.B. Soares Lab  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: [www-bio.llnl.gov/dbp/image/image.html](http://www-bio.llnl.gov/dbp/image/image.html)  
 The following repetitive elements were found in this cDNA sequence:  
 412-461, >HERV571

Seq primer: M13 Forward.  
 Location/Qualifiers  
 source 1..466  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_image="3080089"  
 /clone\_id="NIH-MGC\_50"  
 /tissue\_type="lymph"  
 /cell\_type="germinal center B cells"  
 /cell\_line="MGC85"  
 /lab\_host="DH10B (LTR)"  
 /note="Vector: pT73-Pac; Site\_1: NotI; Site\_2: Eco RI; Constructed from size fractionated cytoplasmic mRNA (3.5-4.4kb). Directionally cloned. Cells provided by Louis M. Staudt, Ph.D. Library preparation by Maria de Fatima Bonaldo, Ph.D. and M. Bento Soares, Ph.D."

BASE COUNT 145 a 90 c 100 g 131 t  
 ORIGIN

Query Match 7.3%; Score 446.8; DB 9; Length 466;  
 Best Local Similarity 97.4%; Pred. No. 1.8e-94;  
 Matches 45; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

OY 4761 CTCGATCCAGACAGTCGCTGCTGCTCCGATTCACCTTAGTCACTAGATCT 4820  
 1 CACGAGCGCAGACAGTCGCTGCTGCTCCGATTCACCTTAGTCACTAGATCT 60  
 Db 1 CACGAGCGCAGACAGTCGCTGCTGCTCCGATTCACCTTAGTCACTAGATCT 60  
 OY 4821 CTCCTCCCAAGTAATGTTCACAGGCCAATTTCAATACCATGTGAGATGTATTT 4880  
 61 CTCCTGCAAAAGTAATGTTCACAGGCCAATTTCAATACCATGTGAGATGTATTT 120  
 Db 61 CTCCTGCAAAAGTAATGTTCACAGGCCAATTTCAATACCATGTGAGATGTATTT 120  
 OY 4881 AATGTATGTATTTAAGAGAACCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4940  
 121 AATGTATGTATTTAAGAGAACCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 180  
 Db 121 AATGTATGTATTTAAGAGAACCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 180  
 OY 4941 GGTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5000  
 181 GGTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 240  
 Db 181 GGTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 240  
 OY 5001 GGGGAAGAAAGAAATTTTAAAAAATTAACATTAAGATGTTTAACTTAAGCTTAA 5060  
 241 GGGGAAGAAAGAAATTTTAAAAAATTAACATTAAGATGTTTAACTTAAGCTTAA 300  
 Db 241 GGGGAAGAAAGAAATTTTAAAAAATTAACATTAAGATGTTTAACTTAAGCTTAA 300  
 OY 5061 TTTGGATGAGAGCAGACACACCGTGACAGCGCTGTATTTTACAGACACACCACTG 5120  
 301 TTTGGATGAGAGCAGACACACCGTGACAGCGCTGTATTTTACAGACACACCACTG 360  
 Db 301 TTTGGATGAGAGCAGACACACCGTGACAGCGCTGTATTTTACAGACACACCACTG 360  
 OY 5121 CTTGAGAGCACAAGTACAGTCGTATCTTAGAAAAGCTTAAAGACCATGTTGAAA 5180  
 361 CTTGAGAGCACAAGTACAGTCGTATCTAGAAAAGCTTAAAGACCATGTTGAAA 420  
 Db 361 CTTGAGAGCACAAGTACAGTCGTATCTAGAAAAGCTTAAAGACCATGTTGAAA 420  
 OY 5181 GAGTCTCAGTTACGAAACAGATGAAGAGCGCTGAGAGGGCT 5226  
 421 GAGTCTCAGTTACGAAACAGATGAAGAGCGCTGAGAGGGCT 466  
 Db 421 GAGTCTCAGTTACGAAACAGATGAAGAGCGCTGAGAGGGCT 466

RESULT 26  
 BF168417 906 bp mRNA linear EST 30-OCT-2000  
 LOCUS BF168417  
 DEFINITION 601776007F1 NCI\_CGAP\_Lu29 Mus musculus cDNA clone IMAGE:4017960 5', mRNA sequence.  
 ACCESSION BF168417  
 VERSION BF168417.1 GI:11048769  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 906)  
 NIH-MGC <http://mgc.nci.nih.gov/>.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
 Tissue Procurement: Gilbert Smith, Ph.D.  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
 Plate: LAM9267 row: m column: 01  
 High quality sequence stop: 638.

FEATURES  
 source 1..906  
 Location/Qualifiers  
 /organism="Mus musculus"  
 /strain="C57BL/6J (f1)"  
 /db\_xref="taxon:10090"  
 /clone\_image="4017960"  
 /clone\_id="NCI\_CGAP\_Lu29"  
 /tissue\_type="spontaneous tumor, metastatic to mammary."  
 /cell\_line="stem cell origin."  
 /lab\_host="DH10B"  
 /note="Organ: Lung; Vector: pCMV-SPORT6; Site\_1: SalI; Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dT. Library constructed by Life Technologies. Investigator providing samples: Gilbert Smith, NIH"

BASE COUNT 242 a 268 c 252 g 144 t  
 ORIGIN

Query Match 7.2%; Score 445.2; DB 10; Length 906;  
 Best Local Similarity 85.4%; Pred. No. 5e-94;  
 Matches 615; Conservative 0; Mismatches 88; Indels 17; Gaps 10;

OY 3579 TCAGGCACAATAGGCCCGAGGAGTATTCCTCCATGCAAGATCCAACTTACACCAT 3638  
 1 TCAGGCACAATAGGCCCGAGGAGTATTCCTCCATGCAAGATCCAACTTACACCAT 59  
 Db 1 TCAGGCACAATAGGCCCGAGGAGTATTCCTCCATGCAAGATCCAACTTACACCAT 59  
 OY 3639 GGGACAGGGGCTATTTAGCCACACTCCGTAATGACGCCAGACGGGCTCAGGCCAC 3698  
 60 GGGACAGGGGCTATTTAGCCACACTCCGTAATGACGCCAGACGGGCTCAGGCCAC 119  
 Db 60 GGGACAGGGGCTATTTAGCCACACTCCGTAATGACGCCAGACGGGCTCAGGCCAC 119  
 OY 3699 GGGCTAGTGCAGAACCAACCAATCAATTAAGACTTCAACTTCAGATGCTTCAAGC 3758  
 120 AGGCAT -GTGAGAACCAACCAACTGAGACTTCAAGCTTCAAGATGCTTCAAGC 178  
 Db 120 AGGCAT -GTGAGAACCAACCAACTGAGACTTCAAGCTTCAAGATGCTTCAAGC 178  
 OY 3759 ACAGGAGAAATGCGCCAGCACTATGATCAATCAATCAAGTATTCATATGCAACTGAC 3818  
 179 ACAGGAGAAATGCGCCAGCACTATGATCAATCAATCAAGTATTCATATGCAACTGAC 237  
 Db 179 ACAGGAGAAATGCGCCAGCACTATGATCAATCAATCAAGTATTCATATGCAACTGAC 237  
 OY 3819 TCTGAGGCTGTGATACCAACAGCAGCAGCTTATATGACACAGATGCTGGCCAGAGACA 3878  
 238 TCTGAGGCTGTGATACCAACAGCAGCAGCTTATATGACACAGATGCTGGCCAGAGACA 297  
 Db 238 TCTGAGGCTGTGATACCAACAGCAGCAGCTTATATGACACAGATGCTGGCCAGAGACA 297  
 OY 3879 GAGGAAATCTGAAACCAACATCTTTCAGACAGACAAATGCAATGCAACAGCAAGTTCA 3938  
 298 GAGGAAATCTGAAACCAACATCTTTCAGACAGACAAATGCAATGCAACAGCAAGTTCA 351  
 Db 298 GAGGAAATCTGAAACCAACATCTTTCAGACAGACAAATGCAATGCAACAGCAAGTTCA 351

OY	3939	GCAAGCACTTTGATGATGAGAGGACAAGGGTGTGAATGATGACACCAGCATGGTGCTCC	3998
Db	3352	GCAGCGACTCTGATGATGATGAGAGGACAGGGCTTGTAATGTGACCCCAGCATGCTGGTCTCC	411
OY	3999	TATGTGATGCCAGCAACTAGTAGSCAACCTCTGGATTCTCCCCAGGCAATGCACGACGTT	4058
Db	412	CGCTGGCCCTTACCAAGCAGCCATGAGCAATCCCGGATTCGCCAAGGCCAAATGGCCAGACGTT	471
OY	4059	TCCATTTCTCCCAACTACAGCAATTAAGTCAGCACCTGATTCACAGCTTACTGGGGCTTC	4118
Db	472	CCCATTTTTCTCTCGAACTACGGAATAAGTCACAAA-CTGATCTCTGGCTTTACTGGGGCTTC	530
OY	4119	GACTCCCAAGAGCCCACTTATGTCAACCCGGAATGGGCATATACACAGATGCCATGATGCA	4178
Db	531	GA-TCCCAAGATCTCTAATGTCTCCCGGATGG-ACATACTCAGAGTCCCATGATGCA	588
OY	4179	ACAGTCTCAGGCGCAACCCAGCCTATCAGGCCCCCTCCCGCATTAATGATGGGCGGAGGG	4238
Db	589	GCAGTCTCAAGCCAACC--AGCTTACCAGGCGCCACTCAGCATGAATGATGGGACCAAGG	646
OY	4239	GAACATGGGCGGAAACAGCATGTTTTCCTCCAGCAGTCCGCCACCACTTTGGGCGACCAAC	4298
Db	647	GAGCATGGGTGG-AACAGCATGTGTGCACAGG--AGTCCACACACACTGTGGAGCAGCAAC	703
<hr/>			
RESULT 27			
LOCUS	AV666871	556 bp	mRNA linear EST 28-NOV-2001
DEFINITION	AV666871 Bos taurus adipocyte cell line Bos taurus cDNA clone		
ACCESSION	E1AD021F12 5', mRNA sequence.		
VERSION	AV666871		
KEYWORDS	AV666871.1 GI:9925901		
SOURCE	EST.		
ORGANISM	cow.		
	Bos taurus		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi;		
	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;		
	Bovidae; Bovinae; Bos.		
	1 (bases 1 to 556)		
	Takasuga,A., Hirotsune,S., Itoh,R., Jitohzono,A., Suzuki,H., Aso,H.		
	and Sugimoto,Y.		
	Establishment of a high throughput EST sequencing system using		
	poly(A) tail-removed cDNA libraries and determination of 36,000		
	bovine ESTs		
	Nucleic Acids Res. 29 (22), E108 (2001)		
JOURNAL MEDLINE	21570554		
COMMENT	Contact: Yoshikazu Sugimoto		
	Animal Genetics Division		
	Shirakawa Institute of Animal Genetics		
	Odakura, Nishigo, Nishi-Shirakawa, Fukushima 961-8061, Japan		
	Tel.: 81-248-25-5641		
	Fax: 81-248-25-5725		
	Email: kazusugi@cocoa.ocn.ne.jp		
	Single pass sequencing.		
	This clone was obtained from a polyA-deleted cDNA library.		
FEATURES	Location/Qualifiers		
SOURCE	1..556		
	/organism="Bos taurus"		
	/db_xref="taxon:9913"		
	/clone="E1AD021F12"		
	/clone_1fb="Bos taurus adipocyte cell line"		
	/cell_type="an adipocyte cell line"		
	/lab_host="DH10B"		
	/note="vector: pZL1, Site_1: SalI; Site_2: NotI; Poly A		
	was deleted from a NotI site"		
BASE COUNT	127 a	186 c	149 g
ORIGIN	94 t		

OY	3331	GGCAATTCTCCAGATGATTTCTATGTCACATCTCCAGCGTAGTCTCCAGATGATCAG	3390
Db	12	GGCACTCTCCCGACGACTTGTCTGTCTCACATCCCGACGAGATCTCCAGTATGAA	71
OY	3391	GGAGCTCTCTGGACACGACTTATCTGGCTTGTGGAAATTTTGTATGCGCTGGAGAGATT	3450
Db	72	GGCGCTCTCTGGACACGACTGATGATGGCTTGGAGAACTTTGATGGCTGGAGAGATC	131
OY	3451	GATAGAGCCTTAGGAATACCCGAAGTGGTCTCAGCCAGAGACCAGAGAGTGTGATCCAGAACAG	3510
Db	132	GACACAGGCCCTGGGGATACCCGAGCTCGTCAAGCCAGAGACCAGAGGTGTGGACCCGAGAGCAG	191
OY	3511	TTTCTAAGTTCAGGATTCACATCATGCTGTGGAGGAGAGGGGCCGTTTCCCAAGCAG	3570
Db	192	TTTCTCAAGTTCAGGATTCACGCTCTATGCTGTAGGACAGAAAGACCCCGCTTTCCCGCAGCAG	251
OY	3571	TATGCAATCTCAGGCGCAAAATGGCCGAGGGTAGCTATTTCTCCATGCAAGATCCAAACTTT	3630
Db	252	TATGGGTCTCAGGGCGCAATGGCCCGAGGGCAGCTACAGCCCATGCAAGATCCAAACTTC	311
OY	3631	CACACCAATGGGACAGCGGCTAGTTATGCCACACTCGGTATGACGCCAGACCAGACCGGGCTC	3690
Db	312	CACACCAATGGGAGACAGCGGCTAGTTATGCCACACTCGGTATGACGCCAGACCAGACCGGGCTC	371
OY	3691	AGGCCACAGGGGCTTAGTGTACAGAACCCAGGCAATATCAATTAAGCTTCAACTTCAGATGCG	3750
Db	372	AGGCCACAGGGGCTTAGTGTACAGAACCCAGGCAATATCAATTAAGCTTCAACTTCAGATGCG	431
OY	3751	CTCCAGACAGCAGAAATGCGCAGCGCACTTATGAATCAAAATCAGCAAGTTTCCAAATGTG	3810
Db	432	CTCCAGACAGCAGCAAAATGCGCAGCGCACTTATGAATCAAAATCAGCAAGTTTCCAAATGTG	491
OY	3811	AACTTGACTCTGAGGCTGTGAGTACCAACACAGGCACTTATTAATGCACAGATGCTGGCC	3870
Db	492	AACTTGACTCTGAGGCTGTGAGTACCAACACAGGCACTTATTAATGAGCCCAATGCTGGCC	551
OY	3871	CAGAG 3875	
Db	552	CAGAG 556	

RESULT	28	
LOCUS	BB658506	
DEFINITION	BB658506 RIKEN full-length enriched, 12 days embryo eyeball Mus musculus cDNA clone D230047M20 5', mRNA sequence.	629 bp mRNA linear EST 26-OCT-2001
ACCESSION	BB658506	
VERSION	BB658506.1	GI:16492331
KEYWORDS	EST.	
SOURCE	house mouse.	
ORGANISM	Mus musculus	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.	
AUTHORS	1 (bases 1 to 629) Arikawa,T., Carinelli,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hiramoto,K., Horii,F., Ishii,Y., Ito,M., Kawai,T., Komno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okazaki,Y., Okita,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.	
TITLE	RIKEN Mouse ESTs (Arikawa,T., et al. 2001)	
JOURNAL	Unpublished (2001)	
COMMENT	Contact: Yoshihide Hayashizaki Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Shuho-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan Tel: 81-45-503-9222 Fax: 81-45-503-9216 Email: genome-res@gs.c.riken.go.jp, URL: http://genome-gsc.riken.go.jp/ Carinelli,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh	

Location/Qualifiers  
1. .629

Query Match	7.28;	Score 443;	DB 9;	Length 629;
-------------	-------	------------	-------	-------------

Matches 497; Conservative 0; Mismatches 70; Indels 2; Gaps 1;

1 GCGGCCGACCTCGGCTACAGCTTCGGCGGCGAAGTCAAGCCGACGGCAGCCGGCA 60

63 GCGGCCGACCTGGGCTCCCGCGGGCGGAGGTCAGCGCCGACGGCAGCCCGCA 122

61 CCTGACGGCGTGACCGACCGAGCCGATTCTCTTGGAATTGGCTACACACTTATAGATC 120

123 CCTGACGGCGTGACGG--CCACATTGATTTCCTCGCATCTGGCTTCACCTGCATTGGCTC 180

121 TTCTGCACTGTTTACAGGCACAGTTGCTGATATGTGTTCAAGATGAGTGGATGGAGAA 180

181 TTCTGCACCTGTGTACAGGCACAGTTGCTGATATGTGTTCAAGATGAGTGGATGGAGAA 240

181 AATACCTCTGACCCCTCCAGGGCAGAGACAAGAAAGCGCAAGGAATGTCCTGACCAACTT 240

241 AACACCTCTGACCCGTCACGAGGAGAGACCAGAAAAACGCAAGGATGTCCCGACCAAGCTC 300

## RESULT 29

LOCUS	BG923396	673 bp	mRNA	linear	EST 05-JUN-2001
-------	----------	--------	------	--------	-----------------

	DEFINITION	002020101.1 NC_016987.1 MMS	MUSCULUS MUSCULUS	CDS	LOC101928000	2
	mRNA sequence.					

```

VERSION      BG923396.1  GI:14303872

```

SOURCE house mouse.

Eukaryota; M

[illegible]

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

COMMENT Contact: Rob

Tissue Procurement: Jeffrey Green M.D.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

Clone distribution: MGC clone distribution information can be.

<http://image.lnl.gov>

High quality sequence stop: 669.

source

BASE COUNT

Query mat

Best Loca







Db 459 TCTTAAAGAACTGTGTAACAGATCCGCCAGATCAAGAGAACAGCCGCTG 400  
 Oy 437 CCAACATAGATGAAGTGCAGAGTGCAGATGTATCTTACAGGGGAGGTGCATTCGCA 496  
 Db 399 CCAACATCGATGAAGTGCAGAGTGCAGATGTATCTTACAGGGGAGGTGCATTCGCA 340  
 Oy 497 AGGATGCGCTGGGCTATGATGCTTGGAGCCCTTGATGGTCTTCTTTAGTAAC 556  
 Db 339 AGGACGACCTGGGCTATGATGCTTGGAGCCCTTGATGGTCTTCTTCTCGTAAC 280  
 Oy 557 TCGAAGCGAACGTTGTCTTGTCTGACAGAAATGACACAGTATCTAAAGTAAACAG 616  
 Db 279 TAGAAGCGAATGTGTCTGCTCCGAGAACGTGACCAATCTGCGGTATTAACAG 220  
 Oy 617 AAGGCTGATGAACAAAGTATATAGCATCTGCATGTTGGGAGCCACACGGAATTTG 676  
 Db 219 AAGACATGATGAACAAAGGCTTTACACATCTGCATGTTGGGAGCCACACGGAATTTG 160  
 Oy 677 TCAAAAACCTGCTCCCAAAGTCTATAGTAATGGGGATCTTGGTCTGGCAACCTCGA 736  
 Db 159 TCAAGAACCTGCTCCCAAAGTCCATGGATGGGGATCTTGGACCGAGAACCTCGCA 100  
 Oy 737 GCGGGAACAGCCATACCTTCAATTTGCGATGCTGTAAACCTTTACCTGATTCAGAG 796  
 Db 99 GCGGGAACAGCCATACCTTCAATTTGCGATGCTGTAAACCTTTGCGATTCAGAG 40  
 Oy 797 AGGAGGCTCATGATTAACGAGAACCTCATCAG 828  
 Db 39 AGGAGGCTCATGATTAACGAGAACCTCATCAG 8  
 RESULT 34 440 bp mRNA linear EST 17-DEC-1999  
 A1654689/c WD49608.x1 NCI\_CGAP\_G66 Homo sapiens cDNA clone IMAGE:2308982.3',  
 LOCUS A1654689 mRNA sequence.  
 A1654689.1 GI:4738668 EST.  
 ORGANISM human.  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 REFERENCE 1 (Bases 1 to 440)  
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgaps-remail.nih.gov  
 Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael  
 R. Emmert-Buck, M.D., Ph.D.  
 cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima  
 Bonaldo, Ph.D.  
 cDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 www.bio.llnl.gov/dbp/ncicgap/ncicgap.html  
 Insert Length: 636 Std Error: 0.00  
 Seq primer: -400P from Glibco  
 High quality sequence stop: 414.  
 location/Qualifiers  
 1.440  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_image="IMAGE:2308982"  
 /clone\_lib="NCI\_CGAP\_G66"  
 /tissue\_type="pooled germ cell tumors"  
 /lab\_host="DH10B"  
 /note="Vector: p773D-Pac (Pharmacia) with a modified  
 polylinker; Site\_1: Not I; Site\_2: Eco RI; Plasmid DNA  
 from the normalized library NCI\_CGAP\_G64 was prepared, and

ss circles were made in vitro. Following HAP purification,  
 this DNA was used as tracer in a subtractive hybridization  
 reaction. The driver was PCR-amplified cDNAs from a pool  
 of 5,000 clones made from the same library (cloneIDs  
 1257096-1259831, 1469064-1470983, and 1475592-1476743).  
 Subtraction by Bento Soares and M. Fatima Bonaldo. "  
 BASE COUNT 144 a 67 c 100 g 128 t 1 others  
 ORIGIN  
 Query Match 6.9%; Score 423.8; DB 9; Length 440;  
 Best local similarity 99.1%; Pred. No. 4.9e-89;  
 Matches 436; Conservative 0; Mismatches 3; Indels 1; Gaps 1;  
 Oy 5096 GCTGATTTACAGACACACCAGTGGCTGAAGACCAACAGTCACTATCTCTAG 5155  
 Db 440 GCTGATTTACAGACACACCAGTGGCTGAAGACCAACAGTCACTATCTCTAG 381  
 Oy 5156 AAAGCTTAAAGACCATGTTGGAAGAGTCCAGTTCTGACAGATGAAGAGAGCT 5215  
 Db 380 AAAGCTTAAAGACCATGTTGGAAGAGTCCAGTTCTGACAGATGAAGAGAGCT 321  
 Oy 5216 GTGAGAGGCTGTAAACATATAGCAAAATATTTTCTGTTTCTTTGTTAAACCA 5275  
 Db 320 GTGAGAGGCTGTAAACATATAGCAAAATATTTTCTGTTTCTTTGTTAAACCA 261  
 Oy 5276 ACTGCTTACCTGATCATGTAATTGAGAGAAATATTTTCAATTTAAAGTCCCT 5335  
 Db 260 ACTGCTTACCTGATCATGTAATTGAGAGAAATATTTTCAATTTAAAGTCCCT 201  
 Oy 5336 TTTAGTTTGAACAGACAGCTGAATCAAGCATCTCTCTCCCTGACGCTGATCTCC 5395  
 Db 200 TTTAGTTTGAACAGACAGCTGAATCAAGCATCTCTCTCCCTGACGCTGATCTCC 141  
 Oy 5396 CTTCCTCTCTCTCAATCCCATCTCTCTATTTCTATCTCTTTTAAATAATA 5454  
 Db 140 CTTCCTCTCTCTCAATCCCATCTCTCTATTTCTATCTCTTTTAAATAATA 81  
 Oy 5455 AGCTACAGAAACAGGTAAGCCCTTATTTCTTAAATGTTTGGCAGCCCTTACCAAT 5514  
 Db 80 AGCTACAGAAACAGGTAAGCCCTTATTTCTTAAATGTTTGGCAGCCCTTACCAAT 21  
 Oy 5515 TGCTAAGTATGAAATTTTCAG 5534  
 Db 20 TGCTAAGTATGAAATTTTCAG 1  
 RESULT 35 580 bp mRNA linear EST 03-APR-2000  
 AM258528 uc36d03.y1 NCI\_CGAP\_Mam5 Mus musculus cDNA clone IMAGE:2811461.5'  
 LOCUS AM258528 similar to TR:061026 Q61026 GRIPL. ;, mRNA sequence.  
 AM258528.1 GI:6631509 EST.  
 ORGANISM house mouse.  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (Bases 1 to 580)  
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgaps-remail.nih.gov  
 Tissue Procurement: Iohar Hennighausen Ph.D., Robin Humphreys  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 www.bio.llnl.gov/dbp/ncicgap/ncicgap.html







QY	302	AAGACCTGCAGAGTGTGATTTTTCGAATTTTAAATGATATGACAACTTAACTCAAC	361
QY	302	AAGACCTGCAGAGTGTGATTTTTCGAATTTTAAATGATATGACAACTTAACTCAAC	361
Db	298	AGGAACCTGCAGAGTGTGATTTTTCGAATTTTAAATGATATGACAACTTCAAC	357
QY	362	CTGACAAATGTGCATCTTAAAGAAAGAACTGTGAGCAAAATTCGTCAGATCAAGAACAG	421
Db	358	CTGACAAATGTGCATCTTAAAGAAAGAACTGTGAGCAAAATTCGTCAGATCAAGAACAG	417
QY	422	AGAAAGCAGCAGCTGCCAACATAGATGAGTGCAGAGTCAAGATCATCTCTACAGGGC	481
Db	418	AGAAAGCAGCAGCTGCCAACATAGATGAGTGCAGAGTCAAGATCATCTCTACAGGGC	477
QY	482	AGGGGTGATGACAAAGGATGCGCGGGGGCCCTATGATCTGTGAGGCGCTTATGGGTCT	541
Db	478	AGGGGTGATGACAAAGGATGCGCGGGGGCCCTATGATCTGTGAGGCGCTTATGGGTCT	537
QY	542	TCTTTGTAGTGAACCT 557	
Db	538	TCTTTGTAGTGAACCT 553	
RESULT 37			
AM027474/c		448 bp	EST 15-SEP-1999
DEFINITION	AM027474	w195.c08.x1 NCI_CGAP_GCG Homo sapiens cDNA clone IMAGE:2515214 3'	
ACCESSION	AM027474	NCI_CGAP_GCG Homo sapiens cDNA clone IMAGE:2515214 3'	
VERSION	AM027474.1	GI:5886230	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	1 (bases 1 to 448)		
TITLE	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.		
JOURNAL	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index		
COMMENT	Unpublished (1997)		
	Contact: Robert Strausberg, Ph.D.		
	Email: cgaps@fmail.nih.gov		
	Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.		
	CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.		
	CDNA Library Arrayed by: Greg Lennon, Ph.D.		
	DNA Sequencing by: Washington University Genome Sequencing Center		
	clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:		
	www.bio.lnl.gov/bdrrp/image/image.html		
	Seq primer: -40up from Gbco.		
FEATURES	Location/Qualifiers		
SOURCE	1. 448		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:2515214"		
	/clone_lib="NCI_CGAP_GCG"		
	/tissue_type="pooled germ cell tumors"		
	/lab_host="DH10b"		
	/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; Plasmid DNA from the normalized library NCI_CGAP_GCG4 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (clones IDs 1257096-1258631, 1469064-1470983, and 1475592-1476743). Subtraction by Bento Soares and M. Fatima Bonaldo."		
BASE COUNT	147 a	69 c	102 g 130 t
ORIGIN			
Query Match	6.88;	Score 420.8;	DB 9; Length 448;

Best Local Similarity   98.9% ; Pred. No.2,5e+88; Indels       2; Gaps       2;					
Matches 445; Conservative   0; Mismatches   2; Indels       3; Gaps       2;					
OY	5086	CGTGCACAGCGCTGATTATTACAGACACACCAGTCGGTAGAGACCACAACAAATCAGTC	5145		
Db	448	CGTGACAGCGGTGATTATTACAGACACACCAGTCGGTAGAGACCACAACAAATCAGTC	389		
OY	5146	GTAATCTAGAAGGCTGAAGAGCCATGTTGGAAAGAGTCCTCACTTACTGAACAGATGA	5205		
Db	388	GTAATCTAGAAGGCTGAAGAGCCATGTTGGAAAGAGTCCTCACTTACTGAAC--ATGA	331		
OY	5206	AAAGAGGCCCTGTGAGAGGCGCTGTAAACAATTAGCAAATATTTTTCTGTTTTCTTTG	5265		
Db	330	AAAGAGGCCCTGTGAGAGGCGCTGTAAACAATTAGCAAATATTTTTCTGTTTTCTTTG	271		
OY	5266	TTAATAACCAACTGGTTACACCTGATTCATGAAATTAG--AAGAAATATTTTCATTTCNA	5324		
Db	270	TTAATAACCAACTGGTTACACCTGATTCATGAAATTAGAAAGAAATATTTTCATTTCNA	211		
OY	5325	ATTAAGTCCCCTTTTGGTTTGGATCACACAGCTTGAAATCAGCATCTTTCCCTGCACG	5384		
Db	210	ATTAAGTCCCCTTTTGGTTTGGATCACACAGCTTGAAATCAGCATCTTTCCCTGCACG	151		
OY	5385	CTGACTCTTCCTCCCTCCCTCTCTCATTTCCCACTCCCTATTTTCATTCCTTTTAAA	5444		
Db	150	CTGACTCTTCCTCCCTCCCTCTCTCATTTCCCACTCCCTATTTTCATTCCTTTTAAA	91		
OY	5445	AAATAATPATAAGCTCAGAGAACACAGTAAAGCCCTTATTTCCCTTAAAGTTTGGCAGC	5504		
Db	90	AAATAATPATAAGCTCAGAGAACACAGTAAAGCCCTTATTTCCCTTAAAGTTTGGCAGC	31		
OY	5505	ACTTACCAATTGCTAAGTATTGAATTTGAG	5534		
Db	30	ACTTACCAATTGCTAAGTATTGAATTTGAG	1		
RESULT	38				
A1393449/c		A1393449	413 bp	mRNA	linear
LOCUS		tg45d08.x1 Soares_NFL_T_GBC_S1	Homo sapiens	cDNA clone	
DEFINITION		IMAGE:2111727 3', mRNA sequence.			
ACCESSION		A1393449			
VERSION		A1393449.1	GI:4222996		
KEYWORDS		EST.			
SOURCE		human.			
ORGANISM		Homo sapiens			
		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
		Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
REFERENCE		1 (bases 1 to 413)			
AUTHORS		NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.			
TITLE		National Cancer Institute, Cancer Genome Anatomy Project (CGAP),			
JOURNAL		Tumor Gene Index			
COMMENT		Unpublished (1997)			
		Contact: Robert Strausberg, Ph.D.			
		Email: cgapbs-re@mail.nih.gov			
		This clone is available royalty-free through LNL ; contact the			
		IMAGE Consortium (info@image.llnl.gov) for further information.			
		Insert Length: 576 Std Error: 0.00			
		Seq primer: -40UP from Glbco			
		High quality sequence stop: 404.			
FEATURES		Location/Qualifiers			
SOURCE		1..413			
		/organism="Homo sapiens"			
		/db_xref="taxon:9606"			
		/clone="IMAGE:2111727"			
		/clone_id="Soares_NFL_T_GBC_S1"			
		/lab_host="DH10B"			
		/note="Organ: pooled; Vector: pTZ19D-Pac (Pharmacia) with			
		a modified polylinker; Site.1: Not I; Site.2: Eco RI;			
		Equal amounts of plasmid DNA from three normalized			
		libraries (fetal lung NDHL19W, testis NHT, and B-cell			
		NCI-GCAP.GCB1) were mixed, and ss circles were made in			
		vitro. Following HAP purification, this DNA was used as			

tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of I.M.A.G.E. clones 297480-302087, 682632-687239, 726408-728711, and 729096-731399. Subtraction by Bento Soares and M. Fatima Bonaldo.

BASE COUNT 139 a 60 c 92 g 122 t  
ORIGIN

Query Match 6.7%; Score 413; DB 9; Length 413;

Best Local Similarity 100.0%; Pred. No. 1.7e-86; Matches 413; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5122 GTGAAGACCAACAAAGTCAGATGCTATCTAGAAAGCTCTAAGACCATGTTGGAAG 5181  
|||||  
Db 413 GTGAAGACCAACAAAGTCAGATGCTATCTAGAAAGCTCTAAGACCATGTTGGAAG 354  
QY 5182 AGTCTCCAGTTACTGAACAGATGAAGAAGACCTGTGAGAGGGCTGTTAACTATGCAAA 5241  
|||||  
Db 353 AGTCTCCAGTTACTGAACAGATGAAGAAGACCTGTGAGAGGGCTGTTAACTATGCAAA 294  
QY 5242 TATTTTTCCTGTTTCTTTCTTTGTTAAACCAACTGCTTACCCTGATCATGTAATGA 5301  
|||||  
Db 293 TATTTTTCCTGTTTCTTTGTTTAAACCAACTGCTTACCCTGATCATGTAATGA 234  
QY 5302 GAAGAATAATTTTCAATTTTCAATTAATTAAGTCCCTTTAGTTGATGATCAGACAGCTTGAAATC 5361  
|||||  
Db 233 GAAGAATAATTTTCAATTTTCAATTAATTAAGTCCCTTTAGTTGATGATCAGACAGCTTGAAATC 174  
QY 5362 AGCATCTCTTCTTCCCTGTCAGACCTGACCTCTCCCTTCCCTCTCTATTTCCCATCTC 5421  
|||||  
Db 173 AGCATCTCTTCTTCCCTGTCAGACCTGACCTCTCCCTTCCCTCTCTATTTCCCATCTC 114  
QY 5422 CCTATTTTCATCTCCTTTTAAATAATATAGTACAGAAACGAGTAAGCCCTTTA 5481  
|||||  
Db 113 CCTATTTTCATCTCCTTTTAAATAATATAGTACAGAAACGAGTAAGCCCTTTA 54  
QY 5482 TTTCCTTAATGTTTTCAGACCACTTACCAATTTGCTAGTATGAAATTTGAG 5534  
|||||  
Db 53 TTTCCTTAATGTTTTCAGACCACTTACCAATTTGCTAGTATGAAATTTGAG 1

RESULT 39

BF651680 485 bp mRNA linear EST 25-Apr-2001  
LOCUS 274527 MARC 3BOV Bos taurus cDNA 5', mRNA sequence.  
DEFINITION BF651680  
ACCESSION BF651680.1 GI:11916810  
VERSION  
KEYWORDS EST.  
SOURCE  
ORGANISM Bos taurus  
Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
Bovidae; Bovinae; Bos.  
1 (bases 1 to 485)  
Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,  
Casas,E., Wray,J.E., White,J., Cho,J., Fahrnenkrug,S.C., Bennett,  
G.L., Heaton,M.P., Laegreid,W.M., Rohrer,G.A., Chitko-McKown,C.G.,  
Perlea,G., Holt,I., Karanycheva,S., Liang,F., Quackenbush,J. and  
Keeler,J.W.

Sequence evaluation of four pooled-tissue normalized bovine cDNA  
libraries and construction of a gene index for cattle  
Genome Res. 11 (4), 626-630 (2001)  
21180013

TITLE

JOURNAL MEDLINE  
COMMENT  
Contact: Smith TPL  
USDA, ARS, US Meat Animal Research Center  
PO Box 166, Clay Center, NE 68933-0166, USA  
Tel: 402 762 4366  
Fax: 402 762 4390  
Email: smith@email.marc.usda.gov

Single pass sequencing. Bases called and alt\_trimmed with phred  
v0.980904.e. Vector identified by cross\_match with the -minscore 18

and -mismatch 12 options.

PCR Primers  
FORWARD: AGGAACACAGTATGACCAT  
BACKWARD: GTTTCACGATCAGCAGC  
Plate: 62 row: A column: 10  
Seq primer: ATTAGGTGACACTATAG.

FEATURES  
source  
1. 485  
/organism="Bos taurus"  
/db\_xref="taxon:9913"  
/clone\_lib="MARC 3BOV"  
/tissue\_type="pooled"  
/lab\_host="DH10B"  
/note="Vector: PCMV SPORT6; Site\_1: XbaI; Site\_2: XhoI;  
Library made from pooled tissue from marrow, alveolar  
macrophage, ovary, fetal semitendinosus muscle, and fetal  
longissimus muscle."

BASE COUNT 125 a 165 c 124 g 71 t  
ORIGIN

Query Match 6.6%; Score 408.8; DB 10; Length 485;

Best Local Similarity 90.3%; Pred. No. 1.7e-85; Matches 437; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 3553 CCGGTTTCCACAGCAGATGATGATCTCAGGCAAAATGGCCCGAGGTACTATTCGCC 3612  
|||||  
Db 2 CCGGTATTTCCACAGCAGATGATGATCTCAGGCAAAATGGCCCGAGGTACTATTCGCC 61  
QY 3613 ATGCAAGATCCAAACTTTCACACCATGAGGACAGCGGCTAGTATGACACATCCGATG 3672  
|||||  
Db 62 ATGCAAGATCCAAACTTTCACACCATGAGGACAGCGGCTAGTATGACACATCCGATG 121  
QY 3673 CAGGCCAGACCGGGGCTCAGAGCCACAGGGGCTAGTGCAGAAACGACCAATCAACTAGA 3732  
|||||  
Db 122 CAGGCCAGACCGGGGCTCAGAGCCACAGGGGCTAGTGCAGAAACGACCCACAGCTGAGG 181  
QY 3733 CTTCACCTTGACATGCGCTTCACAGCAGACAGATGCGCAGCCACTTATGAATCAATC 3792  
|||||  
Db 182 CTTCACCTTGACATGCGCTTCACAGCAGACAGATGCGCAGCCACTTATGAATCAATC 241  
QY 3793 AGCAATGTTTCCAAATGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGAATG 3852  
|||||  
Db 242 AGCAATGTTTCCAAATGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGAATG 301  
QY 3853 AATGACAGATGCTGGCCGACAGACAGAGGGAATCTGTAACAGCATCTTTCAGACAGAGA 3912  
|||||  
Db 302 AATGACAGATGCTGGCCGACAGACAGAGGGAATCTGTAACAGCATCTTTCAGACAGAGA 361  
QY 3913 CAAATGCATCAGCAACAGCAAGTTTCAAGCAAGAACTTGTATGATGAGAGACAGGTTG 3972  
|||||  
Db 362 CAAATGCATCAGCAACAGCAAGTTTCAAGCAAGAACTTGTATGATGAGAGACAGGTTG 421  
QY 3973 AATATGACACACAGCATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4032  
|||||  
Db 422 AATATGACACACAGCATGATGATGATGATGATGATGATGATGATGATGATGATGATG 481  
QY 4033 ATTC 4036  
|||||  
Db 482 CTTC 485

RESULT 40

AI698694/c 410 bp mRNA linear EST 03-JUN-1999  
LOCUS AI698694  
DEFINITION wa80901.x1 Soares\_NFL\_T\_GBC\_Sl Homo sapiens cDNA clone  
IMAGE:2302512 3', mRNA sequence.  
ACCESSION AI698694  
VERSION AI698694.1 GI:4986594  
KEYWORDS EST.  
SOURCE  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;



**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 12, 2002, 03:44:51 : Search time 173.65 seconds  
(without alignments)  
8707.860 Million cell updates/sec

Title: US-09-842-256-1

Perfect score: 6156  
Sequence: 1 GCGGCGCGGACGCGCGCTA.....CCACTTTATGGTGTGTTTT 6156

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database:

Issued\_Patents\_NA:\*  
1: /cgn2\_6/prodata/2/1na/5A\_COMB.seq:\*  
2: /cgn2\_6/prodata/2/1na/5B\_COMB.seq:\*  
3: /cgn2\_6/prodata/2/1na/6A\_COMB.seq:\*  
4: /cgn2\_6/prodata/2/1na/6B\_COMB.seq:\*  
5: /cgn2\_6/prodata/2/1na/PCITUS\_COMB.seq:\*  
6: /cgn2\_6/prodata/2/1na/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6156	100.0	6156	4 US-08-891-640-1	Sequence 1, Appl1
2	6154.4	100.0	6156	4 US-09-723-535-3	Sequence 3, Appl1
3	363.2	5.9	4547	4 US-09-723-534-3	Sequence 3, Appl1
4	58.6	1.0	2082	2 US-08-785-110A-2	Sequence 2, Appl1
5	55.4	0.9	2078	2 US-08-785-110A-1	Sequence 1, Appl1
6	55	0.9	3545	3 US-08-885-291-54	Sequence 54, Appl1
7	55	0.9	3545	4 US-09-496-672-54	Sequence 54, Appl1
8	55	0.9	5715	4 US-09-107-847-1	Sequence 1, Appl1
9	53.8	0.8	7218	1 US-08-232-463-14	Sequence 14, Appl1
10	51	0.8	7218	1 US-08-232-463-14	Sequence 14, Appl1
11	50.2	0.8	4184	2 US-08-785-110A-4	Sequence 4, Appl1
12	48.6	0.8	7498	2 US-08-816-693A-1	Sequence 1, Appl1
13	48.6	0.8	7498	3 US-08-885-291-1	Sequence 1, Appl1
14	48.6	0.8	7498	4 US-09-496-672-1	Sequence 1, Appl1
15	48.2	0.8	2816	4 US-08-785-241-1	Sequence 1, Appl1
16	47	0.8	4010	2 US-08-785-110A-3	Sequence 3, Appl1
17	45.2	0.7	665	2 US-08-883-795A-36	Sequence 36, Appl1
18	43.4	0.7	3031	1 US-08-785-241-2	Sequence 2, Appl1
19	42.6	0.7	1257	1 US-08-640-906-1	Sequence 1, Appl1
20	42.6	0.7	1257	4 US-09-395-936-1	Sequence 1, Appl1
21	42.6	0.7	19124	4 US-08-487-826B-13	Sequence 13, Appl1
22	42.4	0.7	289	4 US-09-007-005-17	Sequence 17, Appl1
23	42.4	0.7	289	4 US-09-244-796-17	Sequence 17, Appl1
24	41.2	0.7	6243	2 US-09-056-075-1	Sequence 1, Appl1
25	41	0.7	1511	1 US-07-991-867B-8	Sequence 8, Appl1
26	41	0.7	1511	1 US-08-107-755A-8	Sequence 8, Appl1
27	41	0.7	1511	2 US-08-544-332-8	Sequence 8, Appl1

C	28	4.1	0.7	19124	2	US-08-487-826B-13	Sequence 13, Appl
C	29	40.8	0.7	2989	3	US-09-068-195-10	Sequence 10, Appl
C	30	40.8	0.7	5565	3	US-09-068-195-21	Sequence 21, Appl
C	31	40.2	0.7	15225	2	US-08-892-403A-2	Sequence 2, Appl
C	32	4.0	0.6	5261	1	US-08-945-806-3	Sequence 3, Appl1
C	33	4.0	0.6	5261	1	US-08-366-051B-3	Sequence 3, Appl1
C	34	39.8	0.6	1000	1	US-08-117-083-63	Sequence 63, Appl
C	35	39.4	0.6	615	4	US-08-998-416-186	Sequence 186, Appl
C	36	39.4	0.6	654	5	PCT-US95-06406A-11	Sequence 11, Appl
C	37	3.9	0.6	6152	4	US-08-973-462-1	Sequence 1, Appl1
C	38	38.6	0.6	1023	1	US-08-256-261-16	Sequence 16, Appl
C	39	38.6	0.6	1023	3	US-08-852-299-16	Sequence 16, Appl
C	40	38.2	0.6	4673	1	US-07-638-431-1	Sequence 1, Appl
C	41	38.2	0.6	4673	5	PCT-US92-00018-1	Sequence 1, Appl1
C	42	3.8	0.6	701	4	US-08-998-416-701	Sequence 701, Appl
C	43	3.8	0.6	246240	4	US-08-724-394A-20	Sequence 20, Appl
C	44	3.8	0.6	246240	2	US-08-724-394A-21	Sequence 21, Appl
C	45	3.8	0.6	246240	2	US-08-724-394A-22	Sequence 22, Appl

#### ALIGNMENTS

RESULT 1  
US-08-891-640-1  
: Sequence 1, Application US/08891640  
: Patent No. 6268173  
: GENERAL INFORMATION:  
: APPLICANT: Chambon, Pierre  
: APPLICANT: Gronemeyer, Hinrich  
: APPLICANT: Voegel, Johannes  
: APPLICANT: Lutz, Yves  
: TITLE OF INVENTION: Transcriptional Intermediary Factor-2  
: NUMBER OF SEQUENCES: 14  
: CORRESPONDENCE ADDRESSES:  
: ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.  
: STREET: 1100 New York Avenue, NW, Suite 600  
: CITY: Washington  
: STATE: DC  
: COUNTRY: USA  
: ZIP: 20005-3934  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Floppy disk  
: COMPUTER: IBM PC compatible  
: OPERATING SYSTEM: PC-DOS/MS-DOS  
: SOFTWARE: Patentin Release #1.0, Version #1.30  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/08/891,640  
: FILING DATE: Herewith  
: CLASSIFICATION: 435  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: US 60/021,247  
: FILING DATE: 12-JUL-1996  
: ATTORNEY/AGENT INFORMATION:  
: NAME: Steffe, Eric K.  
: REGISTRATION NUMBER: 36,688  
: REFERENCE/DOCKET NUMBER: 1383.0130001/EKS  
: TELECOMMUNICATION INFORMATION:  
: TELEPHONE: 202-371-2600  
: TELEFAX: 202-371-2540  
: INFORMATION FOR SEQ ID NO: 1:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 6156 base pairs  
: TYPE: nucleic acid  
: STRANDEDNESS: double  
: TOPOLOGY: linear  
: MOLECULE TYPE: cDNA  
: FEATURE:  
: NAME/KEY: CDS  
: LOCATION: 163..4554  
: US-08-891-640-1

Query Match	Similarity	Score	DB	Length	Matches
Best Local	100.0%	100.0%	Pred. No.	0	0
Matches	6156	Conservative	0	Mismatches	0
			0	Indels	0
			0	Gaps	0
1	GGGGGCGAGGCTGGGCTACAGCTTCGGCGGGAAGGTCAAGCGCCGACGGCGCGCA	60			
1	GGGGGCGAGGCTGGGCTACAGCTTCGGCGGGAAGGTCAAGCGCCGACGGCGCGCA	60			
61	CCTGACGGGGTACGCGACCCGAGCCGATTTCTTGGATTGGCTACACACTTATATATC	120			
61	CCTGACGGGGTACGCGACCCGAGCCGATTTCTTGGATTGGCTACACACTTATATATC	120			
121	TTCTGCACTGTTTACAGGCGACAGTTCCTATATGTGTTCAAGATGAGTGGAGAA	180			
121	TTCTGCACTGTTTACAGGCGACAGTTCCTATATGTGTTCAAGATGAGTGGAGAA	180			
181	AATPACCTGACCCCTCCAGGGGAGAGACAGAAACCGAAGAAATGTCCTGACCAACT	240			
181	AATPACCTGACCCCTCCAGGGGAGAGACAGAAACCGAAGAAATGTCCTGACCAACT	240			
241	GGACCCAGCCCCCAAAAGAACTGAAAAACGTAAATCTGAACAGAGAAATTAATATAT	300			
241	GGACCCAGCCCCCAAAAGAACTGAAAAACGTAAATCTGAACAGAGAAATTAATATAT	300			
301	GAAGAACTTGACAGATGTAATTTTGCAAATTTTAAATGATATAGACAACTTAACTTCAA	360			
301	GAAGAACTTGACAGATGTAATTTTGCAAATTTTAAATGATATAGACAACTTAACTTCAA	360			
361	CCTGACAAATGTGCATCTTTAAAGAACTGTGGAACAAATTCGTGAGTCAAGAACAA	420			
361	CCTGACAAATGTGCATCTTTAAAGAACTGTGGAACAAATTCGTGAGTCAAGAACAA	420			
421	GAGAAAGCAGACAGCTCCAAACATAGATGGAAGTCAGAACTAGATATCTCTACAGG	480			
421	GAGAAAGCAGACAGCTCCAAACATAGATGGAAGTCAGAACTAGATATCTCTACAGG	480			
481	CAGGGTGCATCGACAAAGATGCGGTGGGGCCATGATGCTTGAGGCCCTTGATGGGTTTC	540			
481	CAGGGTGCATCGACAAAGATGCGGTGGGGCCATGATGCTTGAGGCCCTTGATGGGTTTC	540			
541	TTCTTGTATGTAACCTGGAAGGCACGTGTGTGTTGTCAGAGAACTGACACAGTAT	600			
541	TTCTTGTATGTAACCTGGAAGGCACGTGTGTGTTGTCAGAGAACTGACACAGTAT	600			
601	CTPAGGTATACCAAGAAAGCTGATGAAACAAAGTATATAGCATTTGATGTTGGG	660			
601	CTPAGGTATACCAAGAAAGCTGATGAAACAAAGTATATAGCATTTGATGTTGGG	660			
661	GACCAACAGGAATTTGTCAAAAACCTGTGCCAAAGTCTATAGTAAATGGGGGATCTTGG	720			
661	GACCAACAGGAATTTGTCAAAAACCTGTGCCAAAGTCTATAGTAAATGGGGGATCTTGG	720			
721	TCGGGGAACCTCCGAGGGGGAACACCCATACCTTCAATTTGCGAGTGTGTAACCT	780			
721	TCGGGGAACCTCCGAGGGGGAACACCCATACCTTCAATTTGCGAGTGTGTAACCT	780			
781	TTTACGTATTCAGAAAGAGAGGGTCAATGATTAACAGAGAGCTATAGAAATTTGAAGT	840			
781	TTTACGTATTCAGAAAGAGAGGGTCAATGATTAACAGAGAGCTATAGAAATTTGAAGT	840			
841	ATGCAAGTCTTGCGTCTCTCTCAACCAAAAGTCCATCAAAAGAAAGAAAGATTTTGCA	900			
841	ATGCAAGTCTTGCGTCTCTCTCAACCAAAAGTCCATCAAAAGAAAGAAAGATTTTGCA	900			
901	TCCTGCTTATTTGGCTGGCAGAAAGATTTCCATTAAGAAAGACCAATTTCTCCCTCA	960			
901	TCCTGCTTATTTGGCTGGCAGAAAGATTTCCATTAAGAAAGACCAATTTCTCCCTCA	960			
961	TCAGAAAGTTTACTCTCTGCGCAGATTCACAAAGCAAGATCAACCTCTGGATACACG	1020			
961	TCAGAAAGTTTACTCTCTGCGCAGATTCACAAAGCAAGATCAACCTCTGGATACACG	1020			
1021	ACCATGAGGACCACTGAACACCGCTCGGAGGACCTGGTAAAGATGTATTCAGAG	1080			

Db	1021	ACCAATGAGAGCGCATTGAAACCGGCTGGGAGGACCTGTAAGAAGTGTATTGAGAG	1080
Qy	1081	TTCCATGGGAGCAGTATGAAGGAGTAATCGTGTCTTATGCTAAAGAGGCATCATGAACTA	1140
Db	1081	TTCCATGGGAGCAGTATGAAGGAGTAATCGTGTCTTATGCTAAAGAGGCATCATGAACTA	1140
Qy	1141	CTGAGACAAAGATTTGGCATTCAGTCAATCTATGCTTTTCTTGTGTGATGGCACTCTT	1200
Db	1141	CTGAGACAAAGATTTGGCATTCAGTCAATCTATGCTTTTCTTGTGTGATGGCACTCTT	1200
Qy	1201	GTTGTGTGACAAAGAGAGCAAACTATCCGTTCTCAGACTAATTAATGAACCTCAACTT	1260
Db	1201	GTTGTGTGACAAAGAGAGCAAACTATCCGTTCTCAGACTAATTAATGAACCTCAACTT	1260
Qy	1261	GTTAATATCTTTACATATGCTTTTCACAGAGACAAATGTGTGTGATGAATCCGGATCTG	1320
Db	1261	GTTAATATCTTTACATATGCTTTTCACAGAGACAAATGTGTGTGATGAATCCGGATCTG	1320
Qy	1321	ACTGGAACAAGATGGGGAAGCCACTGTAATCAATTAGCTTAACAGCCCTGGCCCATAG	1380
Db	1321	ACTGGAACAAGATGGGGAAGCCACTGTAATCAATTAGCTTAACAGCCCTGGCCCATAG	1380
Qy	1381	GCCCTGTGCTAGTGGGAACCCAGGTGACGAGCATGACCTCAGTAGCAATATTAATTTTCCC	1440
Db	1381	GCCCTGTGCTAGTGGGAACCCAGGTGACGAGCATGACCTCAGTAGCAATATTAATTTTCCC	1440
Qy	1441	ATTAATATGCCCCAAAGGAAACAAATGGGATGCCCATGGGAGGTTTGCTGTGGGGGA	1500
Db	1441	ATTAATATGCCCCAAAGGAAACAAATGGGATGCCCATGGGAGGTTTGCTGTGGGGGA	1500
Qy	1501	ATGAACCAATGTGTACAGCATGCAAGCAACCACTCTCAGGGTAGTAACATGCACTCAAA	1560
Db	1501	ATGAACCAATGTGTACAGCATGCAAGCAACCACTCTCAGGGTAGTAACATGCACTCAAA	1560
Qy	1561	ATGAACAGCCCCCTACAAAGCAGCCCTGGCATTAATCAGAGACAGCCCACTCCATGCTT	1620
Db	1561	ATGAACAGCCCCCTACAAAGCAGCCCTGGCATTAATCAGAGACAGCCCACTCCATGCTT	1620
Qy	1621	TCACCAAGGCATTCGATAGAGCCCTGGAGTGGGAGCCCTCGAATCCCAACCACTGTAG	1680
Db	1621	TCACCAAGGCATTCGATAGAGCCCTGGAGTGGGAGCCCTCGAATCCCAACCACTGTAG	1680
Qy	1681	TTTTTCCCTGCGAGGAGCTTGCAATCCCTCTGTGGGATTTTGGACGACACAGAAATAGC	1740
Db	1681	TTTTTCCCTGCGAGGAGCTTGCAATCCCTCTGTGGGATTTTGGACGACACAGAAATAGC	1740
Qy	1741	CATAGTTATACCAACAGTCCCTTAATGCACTTCAGGCCCTCAGCGAAGGGGACAGGGGTC	1800
Db	1741	CATAGTTATACCAACAGTCCCTTAATGCACTTCAGGCCCTCAGCGAAGGGGACAGGGGTC	1800
Qy	1801	TCATTAAGGGTCATGCTTGGGCTTCACAGACACTTAAATGGGCAATTTGGCAAACTCCCA	1860
Db	1801	TCATTAAGGGTCATGCTTGGGCTTCACAGACACTTAAATGGGCAATTTGGCAAACTCCCA	1860
Qy	1861	GTTAATATGAATTCCTCCCCCACTCAGCAAGANTGGGAAGCTTTGGACTCAAAAGACTGTTT	1920
Db	1861	GTTAATATGAATTCCTCCCCCACTCAGCAAGANTGGGAAGCTTTGGACTCAAAAGACTGTTT	1920
Qy	1921	GGACTATATTTGGGAGCCCTCTGAAGGTACAACTGGACAAAGCAGACAGCTGCGCATCTT	1980
Db	1921	GGACTATATTTGGGAGCCCTCTGAAGGTACAACTGGACAAAGCAGACAGCTGCGCATCTT	1980
Qy	1981	GGAGAGCAAAAGGAAACAAATGACCCCAACCTGCCCCCGGCGTGAGCAGTAGAGACTT	2040
Db	1981	GGAGAGCAAAAGGAAACAAATGACCCCAACCTGCCCCCGGCGTGAGCAGTAGAGACTT	2040
Qy	2041	GACGGGAGAGACAGACTGATGACAGCAAAAGGACAGCAACCAACTCCTGAGCTGCTGAC	2100
Db	2041	GACGGGAGAGACAGACTGATGACAGCAAAAGGACAGCAACCAACTCCTGAGCTGCTGAC	2100
Qy	2101	ACCAAATCTGATCAGATGAGAGCCCTGCGCCTTAGCGAGCTCTTTGTGCGATCAAAACAA	2160

```
Db 2101 ACCAAATCTGATCAGATGAGCCCTGCGCTTACCCAGCTCTTTGTGCGATACAAACAA 2160
QY 2161 GACTCCACAGTAGCTTCCCTGGTTCGTGGCTCTACACATGGAACCTGCTCAAGAGAG 2220
Db 2161 GACTCCACAGTAGCTTCCCTGGTTCGTGGCTCTACACATGGAACCTGCTCAAGAGAG 2220
QY 2221 CATTAATTTTGGACAGACTCTTTCAGAGACAGAGTTCCCTGTGGACTTGGCCAACTTA 2280
Db 2221 CATTAATTTTGGACAGACTCTTTCAGAGACAGAGTTCCCTGTGGACTTGGCCAACTTA 2280
QY 2281 ACAGCAGAACCCACAGCAGCAAGACCTTAGCCAGAGTCCAGACACAGCTCTGTGATCA 2340
Db 2281 ACAGCAGAACCCACAGCAGCAAGACCTTAGCCAGAGTCCAGACACAGCTCTGTGATCA 2340
QY 2341 GAAGTACTATTAAACAAGACCGGTGAGCCCAAGAAGAAAGAGATGCACTACTGCG 2400
Db 2341 GAAGTACTATTAAACAAGACCGGTGAGCCCAAGAAGAAAGAGATGCACTACTGCG 2400
QY 2401 TATTTGCTAGATTAAGATGATACTAAAGATATTGGTTTACCAGAAATTAACCCCAACTT 2460
Db 2401 TATTTGCTAGATTAAGATGATACTAAAGATATTGGTTTACCAGAAATTAACCCCAACTT 2460
QY 2461 GAGGAGATGAGACAGTAAAGACAGATCCCTGCCAGTAAACAAATTTAATAGCAATGAAACT 2520
Db 2461 GAGGAGATGAGACAGTAAAGACAGATCCCTGCCAGTAAACAAATTTAATAGCAATGAAACT 2520
QY 2521 GAGAGAGAGAGATGAGCTTTGAGCTGTGTGACAGCCTGGCAGTGTGAGTGAACACTTG 2580
Db 2521 GAGAGAGAGAGATGAGCTTTGAGCTGTGTGACAGCCTGGCAGTGTGAGTGAACACTTG 2580
QY 2581 GAGAGATTTTGGATGATTTGACAGATAGTCAATTACACAGCTTTTCCAGACAGAG 2640
Db 2581 GAGAGATTTTGGATGATTTGACAGATAGTCAATTACACAGCTTTTCCAGACAGAG 2640
QY 2641 CAGAGGCCCCCTGTGTGATGATGAGTGAACAGCAGCCTCATCATGATGATGCAACTG 2700
Db 2641 CAGAGGCCCCCTGTGTGATGATGAGTGAACAGCAGCCTCATCATGATGATGCAACTG 2700
QY 2701 ACAGCTGAAAAACAGCCCTGTACACACCTGTGTGAGCCAGAAAAAGACAGTGTGCAATTTCA 2760
Db 2701 ACAGCTGAAAAACAGCCCTGTGTACACACCTGTGTGAGCCAGAAAAAGACAGTGTGCAATTTCA 2760
QY 2761 CAGAGCACTTTTAATAAACCCAGCAGCAGGCAACTGGGCACTGTTATGCCAAACAGAAAT 2820
Db 2761 CAGAGCACTTTTAATAAACCCAGCAGCAGGCAACTGGGCACTGTTATGCCAAACAGAAAT 2820
QY 2821 TTACCACTTACATCAGATTCAGATTCAGCAGCAGGAATGATGGTAAATCAAGGATGATA 2880
Db 2821 TTACCACTTACATCAGATTCAGATTCAGCAGCAGGAATGATGGTAAATCAAGGATGATA 2880
QY 2881 AACAGTAGTCCCTACTACAGATGATACCTCAGCCAGGAATGATGGTAAATCAAGGATGATA 2940
Db 2881 AACAGTAGTCCCTACTACAGATGATACCTCAGCCAGGAATGATGGTAAATCAAGGATGATA 2940
QY 2941 GGAACCAAGGAAATTTAGGGAACAGTAGACAGAGGAATGATGGTAAACAGTCTTCGG 3000
Db 2941 GGAACCAAGGAAATTTAGGGAACAGTAGACAGAGGAATGATGGTAAACAGTCTTCGG 3000
QY 3001 CCGACTATGCTATGAGAGATGGCAGCCGACAGATTCGGCTGTGAGAGTCACTGTGCT 3060
Db 3001 CCGACTATGCTATGAGAGATGGCAGCCGACAGATTCGGCTGTGAGAGTCACTGTGCT 3060
QY 3061 GCTACCAACAGTGCATGTAACCGGCGAGTCCAAAGAGATGATGATTCGGAACCCAGCAGCC 3120
Db 3061 GCTACCAACAGTGCATGTAACCGGCGAGTCCAAAGAGATGATGATTCGGAACCCAGCAGCC 3120
QY 3121 AGCATCCCATGAGGCCAGCAGCAGCCGCTGGCCAAAGACAGAGCTTCAGTCTCAGGTC 3180
Db 3121 AGCATCCCATGAGGCCAGCAGCAGCCGCTGGCCAAAGACAGAGCTTCAGTCTCAGGTC 3180
QY 3181 ATGAATATAGGGCATCTGAATTTAGAGATGAAATAGAGGGGACCTCAGTATAGCCAAACA 3240
Db 3181 ATGAATATAGGGCATCTGAATTTAGAGATGAAATAGAGGGGACCTCAGTATAGCCAAACA 3240
QY 3241 CAACTCCCTCCAAATCAGAGTGGCCCATGGGCTGAAGAGATCCCTGCTATAGACAGGCG 3300
Db 3241 CAACTCCCTCCAAATCAGAGTGGCCCATGGGCTGAAGAGATCCCTGCTATAGACAGGCG 3300
QY 3301 TCTTTGCCAGCCAAAACAGGACAGCCATTTGGAGAGTCTCCAGATGATGATGCTATGTCA 3360
Db 3301 TCTTTGCCAGCCAAAACAGGACAGCCATTTGGAGAGTCTCTCCAGATGATGATGCTATGTCA 3360
QY 3361 CATCTCAGCTGAGTCTCCAGATGATGAGGAGCTCTCTGACACAGCTGTATCTGGCC 3420
Db 3361 CATCTCAGCTGAGTCTCCAGATGATGAGGAGCTCTCTGACACAGCTGTATCTGGCC 3420
QY 3421 TTGCGGAATTTGATGAGCTCTGGAGAGATGATGATGAGAGCTTGAAGAAATCCGAACTGGTC 3480
Db 3421 TTGCGGAATTTGATGAGCTCTGGAGAGATGATGATGAGAGCTTGAAGAAATCCGAACTGGTC 3480
QY 3481 AGCCAGAGCCAGAGATGATGATGAGAGAGCTTCAAGTCAAGATTCACATCATGCTG 3540
Db 3481 AGCCAGAGCCAGAGATGATGATGAGAGAGCTTCAAGTCAAGATTCACATCATGCTG 3540
QY 3541 GAGCAGAGAGGCGCCGCTTTTCCACAGCAGTATGATCTCAGGACAAATGGCCAGGGT 3600
Db 3541 GAGCAGAGAGGCGCCGCTTTTCCACAGCAGTATGATCTCAGGACAAATGGCCAGGGT 3600
QY 3601 AGCATATTCCTCCATGACAGATTCCAAACTTTCCACAGAGAGAGAGAGAGAGAGAGAGAG 3660
Db 3601 AGCATATTCCTCCATGACAGATTCCAAACTTTCCACAGAGAGAGAGAGAGAGAGAGAGAG 3660
QY 3661 ACACTCCGATGACAGCCAGACCGGCGCTCAGGCGCCACGCGGCTAGTGAGAAACAGCCA 3720
Db 3661 ACACTCCGATGACAGCCAGACCGGCGCTCAGGCGCCACGCGGCTAGTGAGAAACAGCCA 3720
QY 3721 AATCACTAAGAGCTTCAACTTCAGATCGGCTTCAGACACAGAGAAATGCCAGCCACTT 3780
Db 3721 AATCACTAAGAGCTTCAACTTCAGATCGGCTTCAGACACAGAGAAATGCCAGCCACTT 3780
QY 3781 ATGAATCAAAATCAGCAATGTTTCCAAATGTGAATCTGAGAGCTGTGAGATACCAACA 3840
Db 3781 ATGAATCAAAATCAGCAATGTTTCCAAATGTGAATCTGAGAGCTGTGAGATACCAACA 3840
QY 3841 CAGGACCTATTAATGACAGAGATGCTGCGCCAGAGACAGAGGAAATCCTGAACAGACAT 3900
Db 3841 CAGGACCTATTAATGACAGAGATGCTGCGCCAGAGACAGAGGAAATCCTGAACAGACAT 3900
QY 3901 CTTGAGACAGAGAAATGCAATCAGCAACAGCAAGTTAGCAAGCACTTGTGATGATGAGA 3960
Db 3901 CTTGAGACAGAGAAATGCAATCAGCAACAGCAAGTTAGCAAGCACTTGTGATGATGAGA 3960
QY 3961 GGACAAAGGTTGAATGACAGCAAGATGCTGCTCTAGTGTATGCCAGCAACTATG 4020
Db 3961 GGACAAAGGTTGAATGACAGCAAGATGCTGCTCTAGTGTATGCCAGCAACTATG 4020
QY 4021 AGCAACCCCTGGATTCCTCCAGGCAATGCAACAGTTCATTTCTCCAACTACGGA 4080
Db 4021 AGCAACCCCTGGATTCCTCCAGGCAATGCAACAGTTCATTTCTCCAACTACGGA 4080
QY 4081 ATAAAGTACAGCAACTGATCCAGGCTTTACTGGGGCTACAGCTCCCAAGAGCCCACTTATG 4140
Db 4081 ATAAAGTACAGCAACTGATCCAGGCTTTACTGGGGCTACAGCTCCCAAGAGCCCACTTATG 4140
QY 4141 TCACCCCGAATGGCACATACAGAGATCCCATATGATGACAGTCTCAGGCAACCCAGCC 4200
Db 4141 TCACCCCGAATGGCACATACAGAGATCCCATATGATGACAGTCTCAGGCAACCCAGCC 4200
QY 4201 TATCAGGCCCCCTCCGACATTAATGATGGCGCAGAGGAACTGGCGGAAACAGCATG 4260
Db 4201 TATCAGGCCCCCTCCGACATTAATGATGGCGCAGAGGAACTGGCGGAAACAGCATG 4260
QY 4261 TTTTCCAGAGATGCCACACACTTTGGGCAAGCAAAACACAGCATGTACAGTAAC 4320
Db 4261 TTTTCCAGAGATGCCACACACTTTGGGCAAGCAAAACACAGCATGTACAGTAAC 4320
```

OY	4321	AAACGTAAATCATATGTCCTCATAGGGGACCAACAAGAGGAGGATGACGATGAACAG	4380
OY	4321	AAACGTAAATCATATGTCCTCATAGGGGACCAACAAGAGGAGGATGACGATGAACAG	4380
Db	4321	AAACGTAAATCATATGTCCTCATAGGGGACCAACAAGAGGAGGATGACGATGAACAG	4380
OY	4381	ATGACAGACAGATCAGCATGACCTCAGTACCTCCGTCTACGTCAGGGCTGTCTCC	4440
Db	4381	ATGACAGACAGATCAGCATGACCTCAGTACCTCCGTCTACGTCAGGGCTGTCTCC	4440
OY	4441	ATGGGTCCCGAGCAGGTTAATGATCCTGCTCTGAGGGGAGGCAACCTGTCCAAACAG	4500
OY	4441	ATGGGTCCCGAGCAGGTTAATGATCCTGCTCTGAGGGGAGGCAACCTGTCCAAACAG	4500
Db	4441	ATGGGTCCCGAGCAGGTTAATGATCCTGCTCTGAGGGGAGGCAACCTGTCCAAACAG	4500
OY	4501	CTGCCTGGAAATGATATGATTAAAGCAGGAGGAGACACAACAGGAAATATGTCGACAC	4560
Db	4501	CTGCCTGGAAATGATATGATTAAAGCAGGAGGAGACACAACAGGAAATATGTCGACAC	4560
OY	4561	TGCTGAAGCCAGTTCCTCTTACGTGACCGGGCTCAGTTCCTCAAAACCTCCAGCT	4620
Db	4561	TGCTGAAGCCAGTTCCTCTTACGTGACCGGGCTCAGTTCCTCAAAACCTCCAGCT	4620
OY	4621	GGAGAGCGTGCTATTTGTTTAAACCAACATGACCTGGCAGCCGGTCTGCTAAGAGAG	4680
Db	4621	GGAGAGCGTGCTATTTGTTTAAACCAACATGACCTGGCAGCCGGTCTGCTAAGAGAG	4680
OY	4681	ACAGGCTGGGCTGGTTCCTCCAGGGTGGCTCCACTCGGCTGTGGCAGGAGAGCTGCT	4740
Db	4681	ACAGGCTGGGCTGGTTCCTCCAGGGTGGCTCCACTCGGCTGTGGCAGGAGAGCTGCT	4740
OY	4741	CTTCTCTTGACAGTCTAAGCTGGCATCCAGACAGTCTGCTAGTCTTCCTGCATTCA	4800
Db	4741	CTTCTCTTGACAGTCTAAGCTGGCATCCAGACAGTCTGCTAGTCTTCCTGCATTCA	4800
OY	4801	CCTTAGTCACTAGATCTCTCCCTCCCAAGTAATGTTGACAGCCAAATTCATACC	4860
Db	4801	CCTTAGTCACTAGATCTCTCCCTCCCAAGTAATGTTGACAGCCAAATTCATACC	4860
OY	4861	ATGTCAGATTGATGATATTTAAATGATGATATTAAGAGAAACATGCTCTTGTCCTT	4920
Db	4861	ATGTCAGATTGATGATATTTAAATGATGATATTAAGAGAAACATGCTCTTGTCCTT	4920
OY	4921	CCTGTGGGTTCCAGACAGCTGGTTCCTGCTGTTTCCCGGGCAACAGTCTAGTCC	4980
Db	4921	CCTGTGGGTTCCAGACAGCTGGTTCCTGCTGTTTCCCGGGCAACAGTCTAGTCC	4980
OY	4981	AAAGATTAAAGATTTATCTGGGGAAAGAAAGAAATTTTTTAAAAATTAATTAAGAT	5040
Db	4981	AAAGATTAAAGATTTATCTGGGGAAAGAAAGAAATTTTTTAAAAATTAATTAAGAT	5040
OY	5041	GTTTAAAGCTAAACCTCGAATTTGGGATGGAAGCAGGACAGACCCGTGGACAGCTGT	5100
Db	5041	GTTTAAAGCTAAACCTCGAATTTGGGATGGAAGCAGGACAGACCCGTGGACAGCTGT	5100
OY	5101	ATTTCACAGACACACCAGCTCGTGAAGACCAACAAAGTCAGAGTCTATCTTAAGAC	5160
Db	5101	ATTTCACAGACACACCAGCTCGTGAAGACCAACAAAGTCAGAGTCTATCTTAAGAC	5160
OY	5161	TCTAAAGACATGTTGGAAGAGTCTCCAGTTACTGAAAGATGAAAAGAGCCGTGAG	5220
Db	5161	TCTAAAGACATGTTGGAAGAGTCTCCAGTTACTGAAAGATGAAAAGAGCCGTGAG	5220
OY	5221	AGGCTGTAACTATTACCAAAATATTTTCCCTGTTTTTCTTTGTTAAACCAACAGG	5280
Db	5221	AGGCTGTAACTATTACCAAAATATTTTCCCTGTTTTTCTTTGTTAAACCAACAGG	5280
OY	5281	TTACACCTAAATCATGAAATGGAAGAAATATTTTCAATTCCTAAATTAAGTCCCTTAA	5340
Db	5281	TTACACCTAAATCATGAAATGGAAGAAATATTTTCAATTCCTAAATTAAGTCCCTTAA	5340
OY	5341	TTTGATCAGACAGCTTGAATCAGCATCTCTTCCCTGTACAGCCGACTCTTCCCTTC	5400
Db	5341	TTTGATCAGACAGCTTGAATCAGCATCTCTTCCCTGTACAGCCGACTCTTCCCTTC	5400
OY	5401	CCTGTCATTTCCCAATACCTCCATTTTTCATTTCTTTTAAAAAATATTAAGCTAC	5460

Db	5401	CCTCTCTCATCCCTCCCATACCTCCCTATTTTATTCCTTTTAAAAATATATTAAGTAC	5460
Qy	5461	AGAAACGAGTAAACCCCTTATTTCCCTTAAATTTTGGCAGCACTTACCAATTGCTAA	5520
Db	5461	AGAAACGAGTAAACCCCTTATTTCCCTTAAATTTTGGCAGCACTTACCAATTGCTAA	5520
Qy	5521	GTTATGATTTCCAAAAAATATGATTTACTGGCAGAGGAAAGCAAGTTAAAGC	5580
Db	5521	GTTATGATTTCCAAAAAATATGATTTACTGGCAGAGGAAAGCAAGTTAAAGC	5580
Qy	5581	TTGATACCAATCGAGCTAAAGGATACCGCTTTGGAGCATGTTTATTCGTTCCCCAGCA	5640
Db	5581	TTGATACCAATCGAGCTAAAGGATACCGCTTTGGAGCATGTTTATTCGTTCCCCAGCA	5640
Qy	5641	ACTCGGCCCTCAAAATGGAGAAAGCCAGCTGTGTTAAATGATGAGATATTCAGCA	5700
Db	5641	ACTCGGCCCTCAAAATGGAGAAAGCCAGCTGTGTTAAATGATGAGATATTCAGCA	5700
Qy	5701	CAGATTTAAACCTCGCCATGCTGTTTTTATTTGGTTTTTTAGCAGTCGACTAAGCCGA	5760
Db	5701	CAGATTTAAACCTCGCCATGCTGTTTTTATTTGGTTTTTTAGCAGTCGACTAAGCCGA	5760
Qy	5761	AGTTTGTGAAGTACATAAATCCAAATTTATGTAAACAGCAATATTTAAAGTTGAGA	5820
Db	5761	AGTTTGTGAAGTACATAAATCCAAATTTATGTAAACAGCAATATTTAAAGTTGAGA	5820
Qy	5821	ACTTATGTGTTTAAATGTATTAATTTTGTGAGGTATACATATTTGGAATTCACATAA	5880
Db	5821	ACTTATGTGTTTAAATGTATTAATTTTGTGAGGTATACATATTTGGAATTCACATAA	5880
Qy	5881	AATGAGTACTTGCTGTTTAAATAGATATCTTCATAGCAATGCTCCCAAGGTTT	5940
Db	5881	AATGAGTACTTGCTGTTTAAATAGATATCTTCATAGCAATGCTCCCAAGGTTT	5940
Qy	5941	GTAAGGATATCAATGCGCTTGATTAGACCAATTTGTAGACTTAAGACTTTTATTTCT	6000
Db	5941	GTAAGGATATCAATGCGCTTGATTAGACCAATTTGTAGACTTAAGACTTTTATTTCT	6000
Qy	6001	AAACCTGTGATTTGCTTTATAGTCAATTTATCTATCTATGATGATGACCGCTGTA	6060
Db	6001	AAACCTGTGATTTGCTTTATAGTCAATTTATCTATCTATGATGATGACCGCTGTA	6060
Qy	6061	GGAACCAATTTCTGATTTTATATGTTTATATCTTCTTAATAGAACCTTAGAAGACTA	6120
Db	6061	GGAACCAATTTCTGATTTTATATGTTTATATCTTCTTAATAGAACCTTAGAAGACTA	6120
Qy	6121	CATGTTACTAGCAGGCCACTTTTATGTTGTTTT	6156
Db	6121	CATGTTACTAGCAGGCCACTTTTATGTTGTTTT	6156
RESULT 2			
US-09-723-535-3			
; Sequence 3, Application US/09723535			
; Patent No. 6355483			
; GENERAL INFORMATION:			
; APPLICANT: C. Frank Bennett			
; APPLICANT: Lex M. Cowser			
; TITLE OF INVENTION: ANTISENSE MODULATION OF SRC-2 EXPRESSION			
; FILE REFERENCE: RTS-0225			
; CURRENT APPLICATION NUMBER: US/09/723,535			
; CURRENT FILING DATE: 2000-11-27			
; NUMBER OF SEQ ID NOS: 49			
; SEQ ID NO 3			
; LENGTH: 6156			
; TYPE: DNA			
; ORGANISM: Homo sapiens			
; FEATURE:			
; NAME/KEY: CDS			
; LOCATION: (163)..(4557)			
US-09-723-535-3			



Query Match 100.0%; Score 6154.4; DB 4; Length 6156;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 6155; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGGCGCCGACCCCTGGCTACAGCTTGGCGCGAAGGTGACGCCGACGCCGACGCCGA 60  
Db 1 ggcgcgcgcagcctcggtctacagcttcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 60  
QY 61 CCTGACGCGCTGACCGACGCCGATTTCTTGATTTGGCTTACACTTATAGATC 120  
Db 61 cctgacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 120  
QY 121 TTTCGACATGTTTACAGGACAGTGTGATATGTGTTCAAGATGAGGGATGGGAGAA 180  
Db 121 ttctgcacatgttttacagacagatgtgatatgtgttcaagatgagggatgggagaa 180  
QY 181 AATACTCTGACCCCTCCAGGCGAGACAAAGAAAGCGCAAGGATGCTGACCCACTT 240  
Db 181 aatacctctgacccctccagcgagacaaagaaagcgcaaggatgctgacccaactt 240  
QY 241 GGACCCGACCCCAAGAAAGAACCTGATCTGTGAACAGGAAATTAATATATA 300  
Db 241 ggacccgaccccaagaaagaaacctgatctgtgaacaggaaatataatata 300  
QY 301 GAAGAACCTTGACAGATTGATTTTGCATAATTTAATGATATAGACACTTAACTTCAA 360  
Db 301 gaagaaaccttgacagattgatTTTTGCATAATTTAATGATATAGAACTTAACTTCAA 360  
QY 361 CCTGACAAATGTGCAATCTTAAAGAAACCTGTGAAGCAAAATTCGTGACATCAAGAACAA 420  
Db 361 cctgacaaatgtgcaatctTAAAGAAACCTGTGAAGCAAAATTCGTGACATCAAGAACAA 420  
QY 421 GAGAAACGAGAGCTGCCACATAGATGAAGTGAAGTGAAGTATCTCTTACAGG 480  
Db 421 gagaaacgagagctgccacatagatgaagtgaagtgaagtgaagtgaagtgaagtgaagtga 480  
QY 481 CAGGCTGTCATCGACAGAGATGCGTGGGCTTATGATGCTTGAAGGCCCTTGATGGGTTTC 540  
Db 481 caggctgtcatcgacagagatgCGTGGGCTTATGATGCTTGAAGGCCCTTGATGGGTTTC 540  
QY 541 TTCTTTGATGACCTGGAAGGCAACGTTGTGTTGTGTGACAGAAATGTGACACAGTAT 600  
Db 541 ttctttgatgacctggaaggcaacgTTGTGTTGTGTGACAGAAATGTGACACAGTAT 600  
QY 601 CTAAGGATTAACGAAGAGAGCTGTATGACAAAAGTGTATATGACATTTCCATGTTGGG 660  
Db 601 ctaaggatTAACGAAGAGAGAGCTGTATGACAAAAGTGTATATGACATTTCCATGTTGGG 660  
QY 661 GACCAACGGAATTTGTCAAAACCTGCTGCAAAAGTCTATAGTAAATGGGGGATCTTGG 720  
Db 661 gaccaacggaatTTGTCAAAACCTGCTGCAAAAGTCTATAGTAAATGGGGGATCTTGG 720  
QY 721 TCTGGCAGAACCTCGAGGCGGAGACAGCCATCTTCAATTTGCGATGCTGTAAACCT 780  
Db 721 tctggcagAACCTCGAGGCGGAGACAGCCATCTTCAATTTGCGATGCTGTAAACCT 780  
QY 781 TTACTCTGATTAGAAGAGAGGCTCATGATACAGAGAGCTCATGAAATATGAAACT 840  
Db 781 ttactctgATTAGAAGAGAGGCTCATGATACAGAGAGCTCATGAAATATGAAACT 840  
QY 841 ATGCAGTCTTCGCTGCTCTCAACCAAAATGTCATCAAGAAAGAGAGAGATTTGAG 900  
Db 841 atgcagtCTTCGCTGCTCTCAACCAAAATGTCATCAAGAAAGAGAGAGATTTGAG 900  
QY 901 TCTGCTTGATTTGCTGGCAAGAGATTCCATGAGAGAAAGACAGTTCTTCCCTCA 960  
Db 901 tctgcttgATTTGCTGGCAAGAGATTCCATGAGAGAAAGACAGTTCTTCCCTCA 960  
QY 961 TCACAAGTTTACTACTCGCCAGAGATTCGAAGGACAGATCAGTCTCTGGATACAGC 1020  
Db 961 tcacaagTTTACTACTCGCCAGAGATTCGAAGGACAGATCAGTCTCTGGATACAGC 1020

QY 1021 ACCATGAGAGCCCATGAAACCAAGCTGGAGAGACTGTATGAAGGTATTCAGAG 1080  
Db 1021 accatgagagcccatgAAACCAAGCTGGAGAGACTGTATGAAGGTATTCAGAG 1080  
QY 1081 TTCCATGCGCAGCATGAAGAGAAATCTGTGTCTTATGCTTAAGAGGCATCATGAAATA 1140  
Db 1081 ttccatgCGCAGCATGAAGAGAAATCTGTGTCTTATGCTTAAGAGGCATCATGAAATA 1140  
QY 1141 CTGACACAAAGATTGGCAATTGACATCAATCTGTTTTCCTTGTCTGATGGACCTCTT 1200  
Db 1141 ctgacacAAAGATTGGCAATTGACATCAATCTGTTTTCCTTGTCTGATGGACCTCTT 1200  
QY 1201 GTTGCTGCACAAACGAAGAGCAAACTCATCGCTTTCAGACTACTTAATGAACCTCAACTT 1260  
Db 1201 gttgctgcACAAACGAAGAGCAAACTCATCGCTTTCAGACTACTTAATGAACCTCAACTT 1260  
QY 1261 GTAAATCTTTATCATATGCTTTCACAGAGAGAGAAATGTGTGTGATGAATCCGGATCTG 1320  
Db 1261 gtaaatctTTTATCATATGCTTTCACAGAGAGAGAAATGTGTGTGATGAATCCGGATCTG 1320  
QY 1321 ACTGACAAAGATGGGAGAACCCACTGAATCCATTAATCTGCTTAACAGCCCTGCCATCAG 1380  
Db 1321 actgacAAAGATGGGAGAACCCACTGAATCCATTAATCTGCTTAACAGCCCTGCCATCAG 1380  
QY 1381 GCCCTGTGACATGGGAGAACCCAGTCCAGACATGACCCCTCAGTACGAAATATAATTTTCC 1440  
Db 1381 gccctgtGACATGGGAGAACCCAGTCCAGACATGACCCCTCAGTACGAAATATAATTTTCC 1440  
QY 1441 ATAAATGGCCCAAGAAACAAATGGGATGCCATGGGACAGTTTGTGTTCTGGGGA 1500  
Db 1441 ataaatggCCCAAGAAACAAATGGGATGCCATGGGACAGTTTGTGTTCTGGGGA 1500  
QY 1501 ATGAACCATGTTGTACAGCATGACGATCAAGCAACACTCTCTAGGGTATGATATGCACTCAA 1560  
Db 1501 atgaaccATGTTGTACAGCATGACGATCAAGCAACACTCTCTAGGGTATGATATGCACTCAA 1560  
QY 1561 ATGAACAGCCCTGCAAGAGAGCCCTGTCAGTGAATCCAGAGACCCACCTCCATGCTT 1620  
Db 1561 atgaacAGCCCTGCAAGAGAGCCCTGTCAGTGAATCCAGAGACCCACCTCCATGCTT 1620  
QY 1621 TCACCAAGCATTCGCATGAGCCCTGAGTGGCTGGACAGCCCTCGAATCCACCCAGTCAG 1680  
Db 1621 tcaccaAGCATTCGCATGAGCCCTGAGTGGCTGGACAGCCCTCGAATCCACCCAGTCAG 1680  
QY 1681 TTTTCCCTCGACAGAGACTTTCATTTCCCTGTGGAGATTTTGCAGACGACACAGAAATAGC 1740  
Db 1681 ttttccCTCGACAGAGACTTTCATTTCCCTGTGGAGATTTTGCAGACGACACAGAAATAGC 1740  
QY 1741 CATAGTTTACCAACAGCTCCCTCAATGACATCTCAGGCTCAGGAGAGGGGACAGGGGTC 1800  
Db 1741 catagtTTACCAACAGCTCCCTCAATGACATCTCAGGCTCAGGAGAGGGGACAGGGGTC 1800  
QY 1801 TCATTAGGGTATGTTGGCTTTCACACAGACCTTAAAAAAGGGCAATTTGCCAAAATCTCCCA 1860  
Db 1801 tcattagGGTATGTTGGCTTTCACACAGACCTTAAAAAAGGGCAATTTGCCAAAATCTCCCA 1860  
QY 1861 GTTAATATGATCTCTCCCCCACTCAGCAAGATGGGAGCTTGGACTCAAAAAGACTGTTT 1920  
Db 1861 gttaatATGATCTCTCTCCCCCACTCAGCAAGATGGGAGCTTGGACTCAAAAAGACTGTTT 1920  
QY 1921 GGACTATATGGGAGCCCTTGAAGGTCAACTGGAACAAAGAGAGAGAGAGCTGACATCTT 1980  
Db 1921 ggactATATGGGAGCCCTTGAAGGTCAACTGGAACAAAGAGAGAGAGAGAGCTGACATCTT 1980  
QY 1981 GGAGAGCAAAAGGAAACAAATGACCCCAACCTGCCCCCGGCGTGAGCAGTGAAGAGCT 2040  
Db 1981 ggagagCAAAAGGAAACAAATGACCCCAACCTGCCCCCGGCGTGAGCAGTGAAGAGCT 2040  
QY 2041 GACGGGACAGACAGATCTCATGACAGCAAAAGGACAGACCAATCTCTGACTGCTGAC 2100  
Db 2041 gacggGACAGACAGATCTCATGACAGCAAAAGGACAGACCAATCTCTGACTGCTGAC 2100  
QY 2101 ACCAAATCTGATCAGATGAGACCTCGCCCTTAGCAGACCTTTTGTGAGATCAACAA 2160

|||||  
Db 2101 accaaatccgatcaatgtagcgcctcgcccttagccagctcttctgagatacaaa 2160  
OY 2161 GACTCCACAGTAGCTTGGCTGGTTCTGGTCTACACATGGAACCTCGGTCAAGAGAG 2220  
Db 2161 gactccacagtagcttgcctgcttctggtctacacatggaacccgcctcaaggaag 2220  
OY 2221 CATAAATTTTGCACAGACTCTTGCAGACAGACAGTTCCTGCTGTGACCTGGCCAA 2280  
Db 2221 cataaatcttgacagactcttgacagcagagcttccctctgtgactctgccaagta 2280  
OY 2281 ACAGCGAAGGCCACAGGCGAAAGACCTGAGCCAGAGATCCAGACGACACTCTGATCA 2340  
Db 2281 acagcgaaagccacagcaaaagaccctgagccagagctccagcagcagctctgagta 2340  
OY 2341 GAAGTACATTTAAACAAAGCGCGTGAGCCCCCAAGAAAGAGAGATGACTACTTCCG 2400  
Db 2341 gaagtgactatataaacaagcgggtgagcccaagaagaagaatgcaactactgcg 2400  
OY 2401 TATTGCTGATTAAGATGATACTTAAGATATTGTTTACCGAAATTAACCCCAACTT 2460  
Db 2401 tatttgctagataaagatgataactaaagatattggtttacccaagaataacccccaaact 2460  
OY 2461 GAGAGACTGGACAGTAAAGACAGATCTGCCAGTAAACAAAAATTATAGCAATGAAACT 2520  
Db 2461 gagagactgacagtagaagacagatcctgcagtaacacaaaattaaatgaatgaanaact 2520  
OY 2521 GAGAAAGAGAGATGAGCTTTGAGCTTGAGACACAGCTGGCAGTGAAGTGGAACCTTG 2580  
Db 2521 gaaagaagagagatgagcttggagcctggtgacacagccggcagtgagctggaactctg 2580  
OY 2581 GAGGAGATTTTGGATGATTTGCAAGATAGTCAATTACCAAGCTTTTCCAGACAGAG 2640  
Db 2581 gagagatttgatgattgcatgtaagtagtaacttaaccaagctttccagacagcagag 2640  
OY 2641 CCAGGCGCCCTGCTGATCAGTTGACAAAGCAAGCCATATCATATACCTCAGCAACTC 2700  
Db 2641 ccaggcgccctgctgcatcagltgacaaagcaagcatatacagaccatgcaactc 2700  
OY 2701 ACAGCTGAAAACAGCCCTGCTCAGCTGTGAGCCAGAAAACACAGCTGCGAATTTCA 2760  
Db 2701 acagctgaaaaacagccctgctcaacactgttgagccagaaaaacagcactcggaattca 2760  
OY 2761 CAGAGCACTTTTAAATACCCACGACGAGGCGCAACTGGGCAAGTTATGCCAAACAGAA 2820  
Db 2761 cagagcacttttaataaccacagcagcagggcaactgagcaggttatctgccaacccagaa 2820  
OY 2821 TTACCACTTGACATTCATTTGCAAGCCCAAGCTGGTGGAGACTTTCCACCAATCAGA 2880  
Db 2821 ttaccacttgacatccatgcaaaagcccaacgctgctgagaccttccaccaatcaga 2880  
OY 2881 AACAGTAGTCCCTACTCAGTATGATCTCAGCCAGGATGATGGGTATCAAGGATGATTA 2940  
Db 2881 aacagtagtccctactcactcagtatactctcagccaggaatgataatgaaaggaatgata 2940  
OY 2941 GGAACCAAGAAATTTAGGGAACAGTACGACAGAGATGATGGTAACTGCTTCTCG 3000  
Db 2941 ggaaccaaagaaatttaggaacagtagacaagaaatgattggttaacagtgcttctcgg 3000  
OY 3001 CCACTATGCACTCTGGAAGATGGGCAACGCAAGTTCCGGCTGTGAGAGTCACTGTGCT 3060  
Db 3001 cctactatgccaatctggaagaatgagcagcagagatctggctgtgagagtcacactgtgc 3060  
OY 3061 GCTACACACAGTGCATGGAACCGGCAGTCCAGGAGGTGATTTGGAAACCCAGAGACC 3120  
Db 3061 gctacacacagtgctcatgaaacggccaggtccaaaggaagtatgattggaacccaagacc 3120  
OY 3121 AGCATCCCATGAGGCGCACAGCAGCAGCTGGCCAAAGACAGACGCTTCACTGCTCAGT 3180  
Db 3121 agcatcccatgagggccagcagcagcagcctggccaaagacagcgtctcagtgctcaggtc 3180  
OY 3181 ATGAATATAGGCGCATCTGATTAAGAGATGACATGGGGGACCTCACTATACCAACAA 3240  
Db 3181 atgaatatagggcccatctgaaatagagatgaaatgaggggagcctcagtatagccaaca 3240

Db 3181 atgaatatagggccatctgaaatagagatgaaatgaggggagcctcagtatagccaaca 3240  
OY 3241 CAAGCTCTCTCAAAATCAGACTGCCCCATGCGCTGAAAGCATCTGCTATAGACCAAGCG 3300  
Db 3241 caagctctctcaaatcagacttgccccatgcttgaaagaatctctgctatagaccagggc 3300  
OY 3301 TCTTTTCCAGCCAAAACAGGCAAGCATTTGGCAGTTCTCCAGATGACTTGTATGTCCA 3360  
Db 3301 tcttttccagccaaaacagcagcagcatttgcagcttccacagatgacttgtaagtcca 3360  
OY 3361 CATCTGCAGCTGAGTCTCGATGATGAGGAGGAGCTCTCTGACACAGTGTATCTGGCC 3420  
Db 3361 catctgcagctgagtcctcgatctcgatgagggagctctcttgaccagctgatactgccc 3420  
OY 3421 TTGCGAATTTTGAATGGCTTGAGAGAGATGATGAGCTTGAAGATACCAGACTGTCTC 3480  
Db 3421 ttgcggaatttgatgagccttgagagagatgataagctcttagaatgataccgaactgttc 3480  
OY 3481 AGCCAGGCCAAGCAGTATGATCCAGAACAGTTCTCAAGTCCAGATTTCCAACTCATGCTG 3540  
Db 3481 agccagagccaaagcagtagatccagacagttctcaagtcagagatccacaatcatgtctg 3540  
OY 3541 GAGCAAGAGCGCGCTTTTCCACAGCAGTATGATCTCAGGCAACAAATGGCCAGGGT 3600  
Db 3541 gaggcaagagcgccgctttccacagcagatgatactccagacaaatggccagggc 3600  
OY 3601 AGCTATTCTCCATGCAAGATCCAAACTTTCACACCAGTGGGACAGCGGCTGTTATGCTC 3660  
Db 3601 agctattctccatgaaagatccaacttccacacatgaggaaagggcctagtattgctc 3660  
OY 3661 ACACCTCGTATGACGCCAGACCGGGCTTCAGGCCCACGGGCTTATGCGAAGAACAGCCA 3720  
Db 3661 acactcgtatgacgccagccagccagcggtccagccacagggcctagtgcagaacagcca 3720  
OY 3721 AATCACTTAAGCTTCAACTTCAGATGAGCTTCGCTCCAGAGCAGCAGAGATGCCAGCCACT 3780  
Db 3721 aatcaactaagacttcaacttcaagcatgagcctcccaagcagcagaaatgccagcactt 3780  
OY 3781 ATGAATCAATCAGCAATGTTTCCAAATGTAACCTTGACTGTAGGGCTGGAATCCACACA 3840  
Db 3781 atgaatcaatcagcaatgtttccaatgtaaaccttgactgtagggcctgagttaccaaca 3840  
OY 3841 CAGGCACCTTATTAATGACAGATGCTGGGCCAGAGACAGAGGAATTCCTGAACCGCAT 3900  
Db 3841 caggcaccttataatgacagatgctggccagagacaggaagaaatctctgaacccagcat 3900  
OY 3901 CTTCGACAGAGACAAATGATCAGCAACAGCAAGTTCACCAACGAATTTGATGATGAGA 3960  
Db 3901 cttcgacagagacaaatgcatcagcaaacagcaagttcagcaacgaaacttgatgagaga 3960  
OY 3961 GGAACAAGGTTGAATATGACACCAAGCATGTGGCTCTTAAGTGTATGCCAGCAACTATG 4020  
Db 3961 ggaacaaggttgaaatagacaccaagcaltggtgctcctagtgatgtgcaggaactatg 4020  
OY 4021 AGCAACCTTCGGATTTCCTCCAGGCAATGCAAGCAAGTTTCATTTCTCTCAAACTACGA 4080  
Db 4021 agcaaccttcggatttctccagaggaatgacagcagtttccatttctccaactcagga 4080  
OY 4081 ATTAAGTCAGCAACCTGATCCAGGCTTTACTGGGGCTACGACATCCCGAGGCCCACTTATG 4140  
Db 4081 attaagtcagcaacctgattccaggtcttactggtggctcaggaatcccaagggccaactatg 4140  
OY 4141 TCACCCCGAATGGCAGATACACAGATGCCATGATGATCAACAGTCTCAGGCCAACCCAGCC 4200  
Db 4141 tcaccccgaaatggcagatatacaagagttccatagatgcaaacagctcagccaacccagcc 4200  
OY 4201 TATCAGGCCCCCTCCGACATTAATGATGGGCGCAGGGAACTGGGGCGAAACAGCAATG 4260  
Db 4201 tatcagggccccctccgacataaagtgaatggtcgcaagggaacaatggtgggaacacagcatg 4260  
OY 4261 TTTTCCAGCAGTCCGACACACTTTGGGCAAGCAGCAAAACACAGCATGTAACAATAAC 4320  
Db 4261 ttctccagcagttccccaacacacttggtgagcagcaagcaaacacacagcatgtacagtaac 4320

OY	4321	AAATGAACATCAATGCTCCACATGCGGACCAACACAGCGCATAGACAGATGAACACAG	4380
Db	4321	aaactgaacatcatgctccacatgctgacacaaacaaagctgacatgagacatgaaacag	4380
OY	4381	ATGACAGACAGATGATGACATGACCTCAGTCCGCTGCTTACGTACGTGAGGCTGCTGCC	4440
Db	4381	atgacagagacagatcagcatgacctgaftgacctccgctgcttaagtcagtgaggctgtcc	4440
OY	4441	ATGCGTCCGAGACAGGTTAAATGATCCCTGCTGAGGGAGGACACCTTTGCCAAACAG	4500
Db	4441	atgggtcccgagcaggtttaaigtacctgtctgagggaggaacctgttccaaacag	4500
OY	4501	CTGCGCTGGAATGGATATGATTAAAGCAGGAGGGAGACACACGGAATFATGCTGACAC	4560
Db	4501	ctgctgtgaatgatatgattaaagcagaggagacacacaggaatatgtgtgaac	4560
OY	4561	TGCGAAGCCAGTTGGCTCTTCAGCTGACCGGGCTCAGTGTCTCAAAACATTCCTCAGCT	4620
Db	4561	tgcgaagccagttgtcttccctcagctgacccggctcactgtgtcaaaacatctccagct	4620
OY	4621	GGAGAGCTGTGTTATTTGTTTCAACCCACCACTGACCTGCCACCCGGTCTCTTACAGACAG	4680
Db	4621	ggagagctgtgtattattgttcaaccacactgacctgcacgcoggtctgtctagagcag	4680
OY	4681	ACAGCGCTGCGCCCTGTTCCACAGGAGTGCGCTCAGTCCGCTGTGCGACAGGAGACTGCT	4740
Db	4681	acagcgctgcccgtgttcccaggtgagctccactgtgtgtaggagagctgct	4740
OY	4741	CTTCTCTTGACACTCTGGAAGCTGGCATCCAGACAGTCCGTACGTGTGTTCCTCGATTCA	4800
Db	4741	cttctcttgacagctcgtgaagctgtgcatccagacagctgctgaagtgtgttcttcgtatcca	4800
OY	4801	CCTTAGTGCACCACTTAATCTCTCTCCCACTTAATGTTGACAGGCCCAATTTATATACC	4860
Db	4801	ccttagtgcaactttagatctctctcccaagtaagtgtgacagggccaatttataccc	4860
OY	4861	ATGTGACATTGAAATGATTTAAATGTATGTATTTAAGAGAACCATGCTTGTCTGT	4920
Db	4861	atgtgcattgaaatgattttaaatgtatgtatttaaaggaaccatgtctgttctgt	4920
OY	4921	CCTGTCGGTTCACAGACACTGGTTTCTTTCCTTGTTCCTCGGTAACTGATGATGCC	4980
Db	4921	ccgtgctgggttccagacactggttcttctgtcttgttctccgtgtaacagctcagtgcc	4980
OY	4981	AAAGATTAAAGATTTATCTGGGGGGAAGAAAGAAATTTTAAAAAATTAATCAATAAGAT	5040
Db	4981	aaagattaaagatttatactcgtgggggaagaagaagaattttttaaabaataatacctaagat	5040
OY	5041	GTTTTAAAGCTAAAGCCTGAATTTGGGATGGAAGCAGACACACCGTGAGACAGCGTGT	5100
Db	5041	gttttaagctaaagcctgtaatttggatggaagcaggaacacgcgtgagcagcgctgt	5100
OY	5101	ATTTACGACACACCCAGTCCGCGAAGACCAACAAAGTCACAGTGTGTCTTGAAGAC	5160
Db	5101	atttacgacacacccagctcggtgaagacaaacaaagtacagtgctgtatctctagaagac	5160
OY	5161	TCTTAAACACATTTGGAAGAGTCTCCAGTTACGAAACAGATGAAAGAGACCTGTGAG	5220
Db	5161	tcttaagaccaaigtgtggaagagctcccgattacggaacaggtgaaagagccgtgag	5220
OY	5221	AGGGCGTAAACATTAGCAAAATATTTTTCCTGTGTTTTCCTTTTAAAAACAACATGG	5280
Db	5221	agggcgtttaaacttagcaataatttttccctgttttcttctgtttaaacaacactgtg	5280
OY	5281	TTTCACTGAAATCATGAATTGAGAGAAATATTTTCATTTCTAAATTAAGTCCCTTTAG	5340
Db	5281	tctcacctgaaatcatgattgagaagaataatttcaatttcaaatlaagcccttttag	5340
OY	5341	TTTATATAGACAGTTGAATCAGACATCTCTCTCTGTCGTCAGCTGTGACTCTTCCCTTC	5400
Db	5341	tttatatacagcagcttgaatcagacatcttctctccgtgtaagccgtgacttcttcccttc	5400

QY	5401	CCCTCCCATCTCCCCATACACGCCATTTTCATCTCCCTTTTAAAAAATAATAATAGCTAC	5460
Db	5401	ccctcccatctccccaacacccctatttccattcccttlltaaaaaataataaagctac	5460
QY	5461	AGAAACGAGGTAAACCCCTTTATTTCCCTTAATGTGTTGGCAGCCACTTACCAATGTGTAA	5520
Db	5461	agaaacccagtgaaagcccttatttcccttaaatgttlltgcgcagcaattaccaatgtcaa	5520
QY	5521	GTATTGAATTTTCAGAAAAAAAATGCAATTTACTGCGAAGAGAGCAAGATTAAGGC	5580
Db	5521	gtattgaatttcagaaaaaaaatgcatcttactgccaaggagaagagcaaatgttaagc	5580
QY	5581	TTGATACCAATTCAGAGCTAAAGGATACCTGCTTGGAGAGCATGTTTATCTGTTCCCCACA	5640
Db	5581	ttgataccaatlcgagccaaagafataccgtctlltggaaagcatgttcttccgtccccaagca	5640
QY	5641	ACTCGGCCCTCAAAATGGAGAGAAAGCCGCTGTGTTTAAATGGTAATAGAGATATACAGA	5700
Db	5641	actcgcgcctccaataatggagagaagccagcgtgttctaattgatagcagatatacaga	5700
QY	5701	CAGATTTAAACCTCTGCCATCTGTGTTTATTTTATTTTGGTTTTCAGACAGTGTACATACCGCA	5760
Db	5701	cagatttaacctctgcacatgtgtttttatttgtttttttagcagtgctgactaaagccga	5760
QY	5761	AGTTTGTAGAGTGCATAAAAATCCATTTTATATGTAAACAGCAATTAATTAAGTTGAGA	5820
Db	5761	agtttgtgaagtlacataaataccaaatlctatagtaaaacagcaaatltaagttgaga	5820
QY	5821	ACTTATGTGTTTAAATGTAAATTTTGGAGAGTATACATATGTGGAAATGGACTCAAA	5880
Db	5821	acttatggttctaattgtataatlltcttgagagtaacaatcgtgtgaattgacctcaa	5880
QY	5881	AATGAGTACTTCAGTATTAATTAATTAAGATATCTTCATAGCAATGTCTCTAAAGGTGTTTT	5940
Db	5881	aatgagtgacttcagtatataattagatacttcataagacaatgtctctcctaagtgcttt	5940
QY	5941	GTTAAAGATATCATGCGCTTGATTAGACCTAAATTTGTAACTTAAAGACTTTTATTTCT	6000
Db	5941	gttaagagatatacaatgccttgattagaacctaatlltgbagacttaagacttllattllct	6000
QY	6001	AAACCTTGTGATTTGCTTTTAAAGTCATTTATTTATTCATTAATGTATGATGACAGCCGCTGA	6060
Db	6001	aaacctgtgattctgcttataagtlcaattatcattacatataatgataagcagcgctgta	6060
QY	6061	GGAAACCAATCTTGATTTTATATATGTTTATATTCCTTTTAAATGAGACCTTAGAAAAGCTA	6120
Db	6061	ggaaccaatctctgattcttataatgttataatcttcttctaagaaacttaagaaagacta	6120
QY	6121	CATGTTACTAAGCAGGCACCTTTATAGGTTGTTTTT	6156
Db	6121	catgttactaagcagccacttctaagttgttttt	6156

```

: RESULT 3
US-09-723-534-3
: Sequence 3, Application US/09723534
: Patent No. 6284382
: GENERAL INFORMATION:
: APPLICANT: C. Frank Bennett
: APPLICANT: Lex M. Cowser
: TITLE OF INVENTION: ANTISENSE MODULATION OF SRC-1 EXPRESSION
: FILE REFERENCE: RTS-0225
: CURRENT APPLICATION NUMBER: US/09/723,534
: CURRENT FILING DATE: 2000-11-27
: NUMBER OF SEQ ID NOS: 49
: SEQ ID NO 3
: LENGTH: 4547
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (43)..(4368)
US-09-723-534-3

```



Db 633 CTGGAGGTCCTAGAGCACT 653

```

RESULT 5
US-08-785-310A-1
Sequence 1, Application US/08785310A
Patent No. 5840532
GENERAL INFORMATION:
APPLICANT: McKnight, Steven L.
APPLICANT: Russell, David W.
TITLE OF INVENTION: Neuronal PAS Domain Protein
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
STREET: 268 BUSH STREET, SUITE 3200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/785,310A
FILING DATE: 21-JAN-1997
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: OSMAN, RICHARD A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: UTSO:1226
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 343-4341
TELEFAX: (415) 343-4342
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2078 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-785-310A-1

```

[illegible]

RESULT 6  
US-08-885-291-54  
; Sequence 54, Application US/08885291A  
; Patent No. 6057125  
; GENERAL INFORMATION:  
; APPLICANT: Takahashi, Joseph S.

```

: APPLICANT: Turek, Fred W.
: APPLICANT: Pinto, Lawrence H.
: TITLE OF INVENTION: CLOCK GENE AND GENE PRODUCT
: FILE REFERENCE: 0290-5
: CURRENT APPLICATION NUMBER: US/08/885,291A
: CURRENT FILING DATE: 1997-06-30
: EARLIER APPLICATION NUMBER: 08/816,693
: EARLIER FILING DATE: 1997-03-13
: NUMBER OF SEQ ID NOS: 55
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 54
: LENGTH: 3545
: TYPE: DNA
: ORGANISM: Homo sapiens
: US-08-885-291-54

```

[illegible]

```

RESULT 7
US-09-496-672-54
Sequence 54, Application US/09496672
Patent No. 6291429
GENERAL INFORMATION:
APPLICANT: Takahashi, Joseph S.
APPLICANT: Turek, Fred W.
APPLICANT: Pinto, Lawrence H.
TITLE OF INVENTION: CLOCK GENE AND GENE PRODUCT
FILE REFERENCE: 0290-5
CURRENT APPLICATION NUMBER: US/09/496,672
CURRENT FILING DATE: 2000-02-03
PRIOR APPLICATION NUMBER: 08/885,291
PRIOR FILING DATE: 1997-06-30
PRIOR APPLICATION NUMBER: 08/816,693
PRIOR FILING DATE: 1997-03-13
NUMBER OF SEQ. ID NOS: 55
SOFTWARE: PatentIn Ver. 2.0
SEQ. ID NO 54
LENGTH: 3545
TYPE: DNA
ORGANISM: Homo sapiens
US-09-496-672-54

```

Query Match		0.9%	Score 55;	DB 4;	Length 3545;
Best Local Similarity	55.5%;	Pred.	No. 2.8e-05;		
Matches 106;	Conservative	0;	Mismatches 85;	Indels	0; Gaps
Qy	496	AAGCATGGCGTGGGCGCCATGATGCTGTAGGCCCTTGATGAGTCCTCTTTGTACTGAAC	555		
I					
Dd	735	aatgaagsgtltacaacattaatgtltagagctcctlitgtygtllttlltttgcaatcatg	794		
I					
Qy	556	CTGGAAGGCAACGTTGTGTTTGTGTGCAGAGAATGTGACACAGTATCTAAGGTAAACCA	615		
I					

Db	795	acagatggaagcacaatatatggtctcgagatgtaactcattacttgcacatttccaa	854
QY	616	GAAGAGCGATGAACAAAGGTATATAGCATCTTGATGGGAGCACAGGAATTT	675
Db	855	tcgatctcttgagatcaagaatatttcaatttctccagaagggagactttcagagtc	914
QY	676	GTCAAAAACCT	686
Db	915	tataataatct	925

RESULT 8  
ITS-09-107-

```

1 Sequence 1, Application US/09107847.
2 Patent No. 6100652
3
4 GENERAL INFORMATION:
5
6 APPLICANT: DICKWORTH, DAVID
7 APPLICANT: MICHALOVICH, DAVID
8 TITLE OF INVENTION: NOVEL USE
9 NUMBER OF SEQUENCES: 2
10 CORRESPONDENCE ADDRESS:
11 ADDRESS: Rather 6 Prestia
12 STREET: P.O. Box 980
13 CITY: Valley Forge
14 STATE: PA
15 COUNTRY: USA
16 ZIP: 19482
17
18 COMPUTER READABLE FORM:
19 MEDIUM TYPE: Diskette
20 COMPUTER: IBM Compatidle
21 OPERATING SYSTEM: DOS
22 SOFTWARE: FASTSEQ for Windows Version 2.0
23
24 CURRENT APPLICATION DATA:
25 APPLICATION NUMBER: US/09/107,847
26 FILING DATE: 30-JUN-1998
27
28 CLASSIFICATION:
29
30 PRIOR APPLICATION DATA:
31 APPLICATION NUMBER: EP 97304966.8
32 FILING DATE: 08-JUL-1997
33 ATTORNEY/AGENT INFORMATION:
34 NAME: Prestia, Paul F
35 REGISTRATION NUMBER: 23,031
36 REFERENCE/DOCKET NUMBER: GH-30003
37 TELECOMMUNICATION INFORMATION:
38 TELEPHONE: 610-407-0700
39 TELEFAX: 610-407-0701
40
41 TELEX:
42
43 INFORMATION FOR SEQ ID NO: 1:
44 SEQUENCE CHARACTERISTICS:
45 LENGTH: 5715 base pairs
46 TYPE: nucleic acid
47 STRANDEDNESS: single
48 TOPOLOGY: linear
49
50 MOLECULE TYPE: CDNA
51
52 OS-09-107-847-1

```

Query Match	0.98;	Score 55;	DB 3;	Length 5715;
Best Local Similarity	-55.5%;	Pred. No. 3.9e-05;		
Matches 106; Conservative	0;	Mismatches 85;	Indels 0;	Gaps 0;

[illegible]

QY	676	GTCAAAACCT	686
Db	750	TATAAATACT	760

## RESULT

Sequence 14, Application US/08232463  
Patent No. 5670367  
GENERAL INFORMATION:  
APPLICANT: DORNER, F.  
APPLICANT: SCHEIFLINGER, F.  
APPLICANT: FALKNER, F. G.  
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
NUMBER OF SEQUENCES: 52  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 1800 Diagonal Road, Suite 500  
CITY: Alexandria  
STATE: VA  
COUNTRY: USA  
ZIP: 22313-0299  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/232,463  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/935,313  
FILING DATE:  
APPLICATION NUMBER: EP 91.114 300.6  
FILING DATE: 26-AUG-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703)836-9300  
TELEFAX: (703)683-4109  
TELEX: 899149  
INFORMATION FOR SEQ. ID NO.: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7218 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
CLONE: pTZgpc-Fls  
US-08-232-463-14

Query Match 0.9%; Score 53.8; DB 1; Length 7218;  
Best Local Similarity 8.1%; Pred. No. 0.0001;  
Matches 37; Conservative 225; Mismatches 197; Indels 0; Gaps 0;

[illegible]





Accession	Sequence	Length
D8	TCGTGACAAACAGACGGGACGATCATCTATGTCGTCCGACAGATACACACCTCTCTCTTGGAC	927
QY	ATTAACCAAGAGACCTGATGAAACAAAGTGTATACACTTCGACATGTTGGGACACACA	667
Db	ATTTCACCGCGGATGTCATGTGATCGAATCTGTTAAATTCCTCCAGACCAAGACATTT	987
QY	CGGAATTTGTCAAAACCT	686
Db	CCGAAGTTTATATAATCT	1006

RESULT 12  
ITS-08-816-

```

1 Sequence 1, Application US/0881693A
2 Patent No. 5874241
3
4 GENERAL INFORMATION:
5 APPLICANT: Takahashi, Joseph S
6 APPLICANT: Turek, Fred W
7 APPLICANT: Pinto, Lawrence H
8 TITLE OF INVENTION: Clock Gene and Gene Product
9 NUMBER OF SEQUENCES: 53
10
11 CORRESPONDENCE ADDRESS:
12 ADDRESS: Dressler, Rocky, Milnamow & Katz
13 STREET: Two Prudential Plaza, Suite 4700
14 CITY: Chicago
15 STATE: Illinois
16 COUNTRY: USA
17
18 ZIP: 60601
19
20 COMPUTER READABLE FORM:
21 MEDIUM TYPE: Floppy disk
22 COMPUTER: IBM PC compatible
23 OPERATING SYSTEM: PC-DOS/MS-DOS
24 SOFTWARE: PatentIn Release #1.0, Version #1.30
25
26 CURRENT APPLICATION DATA:
27 APPLICATION NUMBER: US/08/816,693A
28
29 FILING DATE:
30
31 CLASSIFICATION: 435
32 ATTORNEY/AGENT INFORMATION:
33 NAME: No 5874241thnp, Thomas E
34 REGISTRATION NUMBER: 33,268
35 TELECOMMUNICATION INFORMATION:
36 TELEPHONE: 312-616-5400
37 TELEFAX: 312-616-5460
38
39 INFORMATION FOR SEQ ID NO: 1:
40
41 SEQUENCE CHARACTERISTICS:
42 LENGTH: 7498 base pairs
43 TYPE: nucleic acid
44 STRANDEDNESS: single
45 TOPOLOGY: linear
46 MOLECULE TYPE: DNA (genomic)
47
48 FEATURE:
49 NAME/KEY: CDS
50 LOCATION: 389..2954
51
52 US-08-816-693A-1

```

Query Match	Score	DB 2;	Length
0.88;	48.6;	7498;	

Oy	496	AAAGATGGGCGTGGGCGCCATATGATCGTGAAGGCCCTGATGGGCTCTTCTTTGTATGTAAC	555
Db	707	AATGAAGAGTTTACACAGTTAATGTTAGAGGCTCTGATGGTTTTTTTTTATAGCGATATG	766
Oy	556	CTGGAAGCAACGTTGTCTTTGTGTACAGAGATGACACAGTATCTAAGGTATTAACCA	615
Db	767	ACAGATGGAGATATATATATGTATCTGAGGTGTAACCTTGTTACTTGAACATTTACCA	826
Oy	616	GAAGAGCTGATGAACAAAGTATATAGCATCTTGCATGTTGGGGACCAACACGAATTT	675
Db	827	TCTGATCTTTGGAGATCAAGATATATTTAATTTATCCAGAGGGAGAAACATTCAGAGTT	886
Oy	676	GTCAAAAACCT	686

Db 887 TATAAGATACT 897

RESULT 13  
IIS-08-885-

Sequence 1, Application US/08865291A  
Patent No. 6057125  
GENERAL INFORMATION:  
APPLICANT: Takahashi, Joseph S.  
APPLICANT: Turek, Fred W.  
APPLICANT: Pinto, Lawrence H.  
TITLE OF INVENTION: CLOCK GENE AND GENE PRODUCT  
FILE REFERENCE: 0290-5  
CURRENT APPLICATION NUMBER: US/08/885,291A  
CURRENT FILING DATE: 1997-06-30  
EARLIER APPLICATION NUMBER: 08/816,693  
EARLIER FILING DATE: 1997-03-13  
NUMBER OF SEQ ID NOS: 55  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 1  
LENGTH: 7498  
TYPE: DNA  
ORGANISM: Mus musculus  
US-08-885-291-1

Query Match	Score	DB	Length
0.88;	48.6;	3;	7498;

<b>QY</b>	496	A A A N T G G C T G G G G C C T A N G A T G C C T T G A G G C C C T T G A T G G G T T C T T T T T A N G A G A C	555
<b>Db</b>	707	a a t g a a g g t t c a c a c g t t a a t g t a a g g c t c t t g a y g t t c t t t t t t a a g c a t c a y	766
<b>QY</b>	556	C T G A A G G C A A C G T T G T G T T G T C A G A G A N T G C A C A G T A T C T A A G T A T T A A C C A A	615
<b>Db</b>	767	a c a a t g t a a g t a a t a a t a t a t a t g t c t c g a g a g u y t a a c t t c g t a c t t g a a c t t t a c a	826
<b>QY</b>	616	G A A G A C T G A T G A N C A A A G T G T A T A T A G C A T C T T G C A T T T G G G G C C A C A C G A A T T T	675
<b>Db</b>	827	t c t g a t c t t g y g a t c a a g a t a t a t t a a t t t a t t c c c a g a g a g a g a c a t c a a g u t t	886
<b>QY</b>	676	G T C A A A A A A C T T	686
<b>Db</b>	887	t a t a a g a t a c t	897

RESULT 14

```

US-09-496-672-1
: Sequence 1, Application US/09496672
: Patent No. 6291429
:
: GENERAL INFORMATION:
:
: APPLICANT: Takahashi, Joseph S.
:
: APPLICANT: Turex, Fred W.
:
: APPLICANT: Pilito, Lawrence H.
:
: TITLE OF INVENTION: CLOCK GENE AND GENE PRODUCT
:
: FILE REFERENCE: 0290-5
:
: CURRENT APPLICATION NUMBER: US/09/496,672
:
: CURRENT FILING DATE: 2000-02-03
:
: PRIOR APPLICATION NUMBER: 08/885,291
:
: PRIOR FILING DATE: 1997-06-30
:
: PRIOR APPLICATION NUMBER: 08/816,693
:
: PRIOR FILING DATE: 1997-03-13
:
: NUMBER OF SEQ ID NOS: 55
:
: SOFTWARE: Patentln Ver. 2.0
:
: SEQ ID NO 1
:
: LENGTH: 7498
:
: TYPE: DNA
:
: ORGANISM: Mus musculus
:
: US-09-496-672-1

```



[illegible]

```

15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040
1041
1042
1043
1044
1045
1046
104
```

Query Match	0.88;	Score 48.2;	DB 1;	Length 2816;
Best Local Similarity	54.1k;	Pred. No. 0.0023;		
Matches	98;	Conservative	0;	Mismatches 83; Indels 0; Gaps 0;
OY	515	TGAAGCTTGAGCCCTTGATGGCTCTCTCTTTGTAGTAACTGGAGCAACGTTTGCT	574	
Db	418	TGTACCTGAAAGCCCTTGGAGGGTTTCATTCGCCGCTGTACCCCAAGATGGCGACATGATCT	477	
OY	575	TTGTGTGACAGATATGTGACACAGTATCTTAAGGTATTAACCAAGAAAGCGTGGATGAACCAAA	634	

Db	478	TTCTGTGCAAAAACATCAGCAGATTCATGGGACTTAAACACAGTGGAGACTTAACGAGACATA	537
Oy	635	GTGTAATATAGCATCTTGCATGTTGGGACACACGAGATTTGTCAAAAACCTGCTGCCAA	694
Db	538	GTACTCTTGACTTCATCATTCCTCCGACACCATGAGAGATTCTGAGAAACCTGAGTCTCA	597
Oy	695	A 695	
Db	598	A 598	

```

: RESULT 16
: US-08-785-310A-3
: Sequence 3, Application US/08785310A
: Patent No. 5840532
: GENERAL INFORMATION:
: APPLICANT: McKnight, Steven L.
: APPLICANT: Russell, David W.
: TITLE OF INVENTION: Neuronal PAS Domain Protein
: NUMBER OF SEQUENCES: 8
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
: STREET: 268 BUSH STREET, SUITE 3200
: CITY: SAN FRANCISCO
: STATE: CALIFORNIA
: COUNTRY: USA
: ZIP: 94104
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/785,310A
: FILING DATE: 21-JAN-1997
: CLASSIFICATION: 536
: ATTORNEY/AGENT INFORMATION:
: NAME: OSMAN, RICHARD A
: REGISTRATION NUMBER: 36,627
: REFERENCE/DOCKET NUMBER: UTSID.1226
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 343-4341
: TELEFAX: (415) 343-4342
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 4010 base pairs
: type: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: CDNA
: IS-08-785-310A-3

```

Query Match	0.88;	Score 47;	DB 2;	Length 4010;
Best Local Similarity	52.38;	Pred. No. 0.0065;		
Matches 104;	Conservative 0;	Mismatches 95;	Indels 0;	Gaps 0;
QY 488 TCATCGACAAAGATGCGCTGGGGCCCTATGATGCTTGAGGCCCTTGATGATGGTTCTTCTTTG 547				
Db 522 TCCTCGATTAATGAGAAATTCACCCAGCGTGAAGTTGGAGGCAATTAGATGGCTTCATTAATCG 581				
QY 548 TAGTGACCTCGAAGGCGCAACTGTGTGTTGTCTGAGAGAAATGTGACACAGTATCTTAAGT 607				
Db 582 CAGTGACAAACAGAGCGGCGCATCACTATATGTCTCGAACAGTATCAAGCTCTCTCTTGGAC 641				
QY 608 ATTAACCAAGAGACCTGATGTAACAAAAAGTATATATGACATCTTTCGATTTGGGACGCACA 667				
Db 642 ATTATACCGTGGGATGTATGATCATGAGATTTGTTAAATTTCTCCCAACAAGAACATTT 701				
QY 668 CGGATTTGTCAAAAACCT 686				
Db 702 CAGAGATTTATATAAATCCT 720				

RESULT 17  
US-08-883-795A-36  
; Sequence 36, Application US/08883795A  
; Patent No. 5985607  
; GENERAL INFORMATION:  
; APPLICANT: Delcave, Genevieve  
; APPLICANT: Awang, Gregor  
; TITLE OF INVENTION: Recombinant DNA Molecules and Expression  
; TITLE OF INVENTION: Vectors for Tissue Plasminogen Activator  
; NUMBER OF SEQUENCES: 39  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BERSKIN & PARR  
; STREET: 40 King Street West  
; CITY: Toronto  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: M5H 3Y2  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/883,795A  
; FILING DATE: 27-JUN-1997  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Gravelle, Micheline  
; REGISTRATION NUMBER: 40,261  
; REFERENCE/DOCKET NUMBER: 7841-062  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (416) 364-7311  
; TELEFAX: (416) 361-1398  
; INFORMATION FOR SEQ ID NO: 36:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 665 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; ORIGINAL SOURCE:  
; ORGANISM: Homo sapiens  
; IMMEDIATE SOURCE:  
; CLONE: Rh 32  
; US-08-883-795A-36

Query Match 0.7%; Score 45.2; DB 2; Length 665;  
Best Local Similarity 47.0%; Pred. No. 0.0061;  
Matches 140; Conservative 0; Mismatches 158; Indels 0; Gaps 0;  
QY 5796 AAACAACATATTTAAGTGAAGAACTTATGCTTTTATTTATTTATTTTGTGAGGT 5855  
DB 3 AAAGATGAGGTAATGTGTTTATTAATTAATTAATTAATTAATTAATTAATTAATTA 62  
QY 5856 ATACATATTTGGAATTAAGTCAAAATGAGTACTTCAGTATTAATTAATTAATTAATTA 5915  
DB 63 AAATATTTTAAATTAATTAATTTATTAATTAATTAATTAATTAATTAATTAATTAATTA 122  
QY 5916 TACCAATGTCCTAAAGGTGTTTGAAGATATCAATGCTTGATTAAGACCTAATTT 5975  
DB 123 AAATATTTTAAATTAATTAATTTATTAATTAATTAATTAATTAATTAATTAATTAATTA 182  
QY 5976 GTAGACTTAAGACTTTTATTTTCTAAACCTTGATTCGTTATTAAGACATTTATCTA 6035  
DB 183 AAATATTTTAAATTAATTAATTTATTAATTAATTAATTAATTAATTAATTAATTAATTA 242  
QY 6036 ATCTATATGATATGACGCCCTGTAGGAACCAATTCGTGATTTTAAATTTTAAATTT 6093  
DB 243 AAATATTTTAAATTAATTAATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 300

RESULT 18  
US-08-785-241-2  
; Sequence 2, Application US/08785241  
; Patent No. 5655963  
; GENERAL INFORMATION:  
; APPLICANT: McKnight, Steven L.  
; APPLICANT: Russell, David W.  
; APPLICANT: Tian, Hui  
; TITLE OF INVENTION: Endothelial PAS Domain Protein  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP  
; STREET: 268 BUSH STREET, SUITE 3200  
; CITY: SAN FRANCISCO  
; STATE: CALIFORNIA  
; COUNTRY: USA  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/785,241  
; FILING DATE: 17-JAN-1997  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: OSMAN, RICHARD A  
; REGISTRATION NUMBER: 36,627  
; REFERENCE/DOCKET NUMBER: UTSD:1229  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 343-4341  
; TELEFAX: (415) 343-4342  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3031 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; US-08-785-241-2

Query Match 0.7%; Score 43.4; DB 1; Length 3031;  
Best Local Similarity 52.5%; Pred. No. 0.059;  
Matches 95; Conservative 0; Mismatches 86; Indels 0; Gaps 0;  
QY 515 TGATGCTTGAGGCCCTTGATGAGTCTTCTTTGTAAGCACTGGAAGCAACCTTGCT 574  
DB 368 TGACCTGAAGAGCCCTTGAGAGGTTTCATTCGTGTGACCAAGAGGATGATGATCT 427  
QY 575 TTGTGACAGAGATGAGACACATATCTAAGTATTAACCAAGAGAGCGATGAACAAA 634  
DB 428 TTCTGTGGAAACATCAGCAAGTTCATGAGACTTACTCAGGTAGACTTAACGACACA 487  
QY 635 GTGTATATAGCATCTTGACATGTTGGGACCAACAGAAATTTGCAAAAACCTGTCCAA 694  
DB 488 GCATCTTGACTTCACATCAATCCCTGGGACCATGAAGATCCGTGAAACCTGACTCTCA 547  
QY 695 A 695  
DB 548 A 548

RESULT 19  
US-08-640-906-1  
; Sequence 1, Application US/08640906B  
; Patent No. 6140100  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Gary K  
; APPLICANT: Blumenkopf, Todd A.  
; APPLICANT: Cory, Michael





MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/056,075  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Seay, Nicholas J.  
REGISTRATION NUMBER: 27386  
REFERENCE/DOCKET NUMBER: 960296,95238  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 608-251-5000  
TELEFAX: 608-251-9166  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 6243 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: misc.feature  
LOCATION: 3770..4013  
OTHER INFORMATION: /note="Rp4 origin of DNA transfer (orig)" from  
OTHER INFORMATION: plasmid Rp4"  
US-09-056-075-1

Query Match 0.7%; Score 41.2; DB 2; Length 6243;  
Best Local Similarity 47.5%; Pred. No. 0.43;  
Matches 153; Conservative 0; Mismatches 168; Indels 1; Gaps 1;

QY 5774 ACATFAAAATCCAAATTTATGTAACAAGCAATTAATTAAGTGAAGAACTATGTTT 5833  
DB 1362 AAAATFAAAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1421  
QY 5834 AATTGTAAATTTTGTAGAGTATACATATGTGGAATTTGCTCAAAATAGACTTC 5893  
DB 1422 TTGAAAAAATTTTATATATATATA-ATCTTGAAAGAAAAGAAATATAAAGGACC 1480  
QY 5894 AGTATTAATAGATATCTCATAGCAATGCTCCCTAAGGTTTGTAAAGATATCA 5953  
DB 1481 TTTTAAAGCCATTTTTCATATACGTAATATGCGCTTCAATTTTATGTA 1540  
QY 5954 ATGCCCTGATTAGACCTAATTTGAGACTTAAAGCTTTTATTTCTAAACCTGTGAT 6013  
DB 1541 CTCTACATTAAGATTAATTTCTTATTTTAAAGCCCTTTCTTAAAGGCTTTATTT 1600  
QY 6014 CTGCTTAATCACTATATATCAATCAATATATGATATGACGCGCTGTAAGCAATTC 6073  
DB 1601 TTTTCTTAATACATTAATTAATCTTTTGTGCTTTTCCCTTAAGCTTTTATTTCT 1660  
QY 6074 GATTTTATATGTTTATATCT 6095  
DB 1661 CTGATTAATTTTCTTACCTCT 1682

RESULT 25  
US-07-991-867B-8/C  
Sequence 8, Application US/07991867B  
Patent No. 5476781  
GENERAL INFORMATION:  
APPLICANT: Moyer, Richard W.  
APPLICANT: Hall, Richard L.  
APPLICANT: Gruidl, Michael E.  
TITLE OF INVENTION: No. 5476781el Entomopoxvirus Expression System  
NUMBER OF SEQUENCES: 66  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: David R. Saliwanchik  
STREET: 2421 N.W. 41st Street, Suite A-1  
CITY: Gainesville

STATE: FL  
COUNTRY: USA  
ZIP: 32606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/991,867B  
FILING DATE: 12-DEC-1992  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO 92/14818  
FILING DATE: 12-FEB-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/827,685  
FILING DATE: 30-JAN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/657,584  
FILING DATE: 19-FEB-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Saliwanchik, David R.  
REGISTRATION NUMBER: 31,794  
REFERENCE/DOCKET NUMBER: UP114.C3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 904-375-8100  
TELEFAX: 904-372-5800  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1511 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:  
ORGANISM: Amsacta moorei entomopoxvirus  
FEATURE:  
NAME/KEY: CDS  
LOCATION: complement (18..218)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: complement (234..782)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 852..1511  
US-07-991-867B-8

Query Match 0.7%; Score 41; DB 1; Length 1511;  
Best Local Similarity 47.0%; Pred. No. 0.18;  
Matches 165; Conservative 0; Mismatches 180; Indels 6; Gaps 1;

QY 5773 TACATFAAAATCCAAATTTATGTAACAAGCAATTAATTAAGTGAAGAACTATGTTT 5832  
DB 1143 TACTATATATGCTATTTTGTATATAGATATATCTAAATATATATTTTATTTG 1084  
QY 5833 TAAITGTAAATTTTGTAGAGTATACATATTTGGGAATTTGACATCAAAATAGACTACT 5892  
DB 1083 TTATFAAAATTTAAATTAATTAATTAATTTGAATATATTAATTAATTAATTTCTG 1024  
QY 5893 CAGATTAATTAAGATATCTCATAGCAATGCTCCCTAAGGTTTGTAAAGATATTC 5952  
DB 1023 GAAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 964  
QY 5953 AATGCCCTGATTAGACCTAATTTGAGACTTAAAGCTTTTATTTCTAAACCTGTGAT 6012  
DB 963 ATTTTAAATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 910  
QY 6013 TCTGCTTAATGCTATTAATCTATATGATATGACGCGCTGTAAGCAATTC 6072  
DB 909 TTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 850

OY 6073 TGAATTTTATATCTTATATCTTCTTAATGACCTTAGAAGACTACAT 6123  
DB 849 TGAATATTTATTTTCTTTCATGATTAATTTTTCGAAAAATATAT 799

## RESULT 26

US-08-107-755A-8/C  
Sequence 8, Application US/08107755A  
Patent No. 5721352  
GENERAL INFORMATION:  
APPLICANT: Moyer, Richard W.  
APPLICANT: Hall, Richard L.  
APPLICANT: Gruidl, Michael E.  
TITLE OF INVENTION: No. 5721352e1 Entomopoxvirus Expression System  
NUMBER OF SEQUENCES: 40  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: David R. Saliwanchik  
STREET: 2421 N.W. 41st Street, Suite A-1  
CITY: Gainesville  
STATE: Florida  
COUNTRY: U.S.A.  
ZIP: 32606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/107,755A  
FILING DATE: 19-AUG-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/837,658  
FILING DATE: 30-JAN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/657,584  
FILING DATE: 19-FEB-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Saliwanchik, David R.  
REGISTRATION/DOCKET NUMBER: 31,794  
REFERENCE/DOCKET NUMBER: UF114.C2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (904) 375-8100  
TELEFAX: (904) 372-5800  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1511 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:  
ORGANISM: Amsacta moorei entemopoxvirus  
FEATURE:  
NAME/KEY: CDS  
LOCATION: complement (18..218)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: complement (234..782)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 852..1511  
US-08-107-755A-8

Query Match 0.7%; Score 41; DB 1; Length 1511;

Best Local Similarity 47.0%; Pred.No.0.18; Mismatches 180; Indels 6; Gaps 1;

OY 5773 TACATAAATCCATTTATATGTAACAAGCATTATTTAGTTGAGACTATGTTT 5832  
DB 1143 TACTTATATTCATTTTCTTATTAAGATATATCTAAATATGTTATATTTTAAATTTTG 1084

OY 5833 TAATTTATATATTTTGTGAGATATACATATTTGCAATTTAGCTCAAAAATGAGTACTT 5892  
DB 1083 TATATAAATTTAAATTAATTAATTAATTTAAATTTGAAATATATAAATTTAAATTTCTG 1024  
OY 5893 CAGTATTAATTAATGATATCTTCATGCAATGCTCCTAAAGGCTTTGTAAGGATATC 5952  
DB 1023 GAATATTTATTAATAATATATTTATCTATATATTTATATGCAATTTCTTTAAATTTACTA 964  
OY 5953 AATGCTTATTAAGCTATTAATTTGTAAGCTTTTAATTTTCTAAACCTTGAT 6012  
DB 963 ATTTTATTAATATTAATATTAATTAATTAATTAATTTATCTGTTTATTTT-----ATATTTT 910  
OY 6013 TCTGCTTATTAAGCTATTTATCTATATATGATATGACGCGCTTAGAAGCAATTTCT 6072  
DB 909 TTAATTAATTAATTTTAAATTTATTTATTAATTAATTAATTAATTAATTAATTAATTTT 850  
OY 6073 TGAATTTTATGTTATATCTTCTTCTTAATGACCTTAGAAGACTACAT 6123  
DB 849 TGAATATTTATTTTCTTTCATGATTAATTTTTCGAAAAATATAT 799

## RESULT 27

US-08-544-332-8/C  
Sequence 8, Application US/08544332  
Patent No. 5935777  
GENERAL INFORMATION:  
APPLICANT: Moyer, Richard W.  
APPLICANT: Hall, Richard L.  
APPLICANT: Gruidl, Michael E.  
TITLE OF INVENTION: No. 5935777e1 Entomopoxvirus Expression System  
NUMBER OF SEQUENCES: 77  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Gerard H. Bencen  
STREET: 2421 N.W. 41st Street, Suite A-1  
CITY: Gainesville  
STATE: FL  
COUNTRY: USA  
ZIP: 32606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/544,332  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/991,867  
FILING DATE: 07-DEC-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/107,755  
FILING DATE: 19-AUG-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO 92/14818  
FILING DATE: 12-FEB-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/837,685  
FILING DATE: 30-JAN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/657,584  
FILING DATE: 19-FEB-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Bencen, Gerard H.  
REGISTRATION/DOCKET NUMBER: 35,746  
REFERENCE/DOCKET NUMBER: UF114.C4  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 904-375-8100  
TELEFAX: 904-372-5800  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1511 base pairs  
TYPE: nucleic acid







Query Match	0.7% ;	Score 40.2 ;	DB 2 ;	length 15225 ;
Best Local Similarity	45.28 ;	Pred. No. 1.6 ;		
Matches 147 ;	Conservative 0 ;	Mismatches 178 ;	Indels 0 ;	Gaps 0 ;
QY 5808	ATTTAACTTGGAACCTTATGTCGTTTTAAATGTATATATTTGTGAGGATACATATTTGTG			5867
Db 13754	ATTTATTTAGGAATGTTGACATCATTTTAAATTCGGATTTCCCTTATACAAATTTTA			13695
QY 5868	GAATTGACTCAAAAATGAGTACTTCAGTATTATTAATTTGATATCTTCATAGCAATGTC			5927
Db 13694	GGTTTGTACTATATATATCTTTAAACAGCAATTAATGATATATATTTTCTAAAGTTTCTGG			13635
QY 5928	CTAAGGCTGTTTGTAAAGATATCAATCAAGCCCTGTTGTATAGACTAATTTGTGACTTAGA			5987
Db 13634	GTTGGGTGATATAGTTGTTTATATATATATATCTCTATATCTGAATTAGCAATCTTATTTG			13575
QY 5988	CTTTTATTTTCTAAACCTTGATTCGCTTATATAGCATTTATCTATATATGATA			6047

```

RESULT 32
US-08-045-806--3
; Sequence 3, Application US/08045806
; Patent No. 5378822
; GENERAL INFORMATION:
; APPLICANT: Bradfield, Christopher Alan
; APPLICANT: Dolwick, Kristin Marie
; APPLICANT: Poland, Alan
; TITLE OF INVENTION: Ah Receptor cDNA and Method of
; TITLE OF INVENTION: Determining Human Risks To Environmental Pollutants
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Tilton, Fallon, Lungmus & Chestnut
; STREET: 100 South Wacker Drive, Suite 960
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606-4002
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/045,806
; FILING DATE: 19930408
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Pentress, Susan B.
; REGISTRATION NUMBER: 31,327
; REFERENCE/DOCKET NUMBER: NU-9207
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312)-456-8000
; TELEFAX: (312)-456-7776
; INFORMATION FOR SEQ. ID NO.: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5261 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 383..2927
; US-08-045-806--3

```

	Query Match	Best Local Similarity	Score 40;	DB 1;	Length 5261;
	Matches 82;	Conservative	0.6%;	53.9%;	Pred. No. 0.85;
			0;	Mismatches 70;	Indels 0;
					Gaps 0;
QY	524	AGGCCCTTGATGGGGCTTCCTCTTTGTAGTGAACCTCGAAGCAACCTTGCTTTGTGTGCAG	583		
Db	741	AGGCTCTGAATGGCTTTGTATTATTAGTTGCACACACAGATGCTTTGGCTTTTATGCTTCTT	800		
QY	584	AGATGTGACACAGATATCTTAAGTATTAACCAAGAAGAGCTGATGGAACAAAGCTATATATA	643		
Db	801	CTACATATACAGATATATCTAGGGTTTCAGACAGTGTGATGTCTATCATCATCAGACTATATATG	860		
QY	644	GCATCTTCGATCTTGGGGACCAACAGGAAATTT	675		
Db	861	AACCTTATCCATACCGAAGAGCCGAGCTGAATTT	892		



STREET: 3054 Cornwallis Road  
CITY: Research Triangle Park  
STATE: No. 6239264th Carolina  
COUNTRY: USA  
ZIP: 27709  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/998,416  
FILING DATE: 24-DEC-1997  
CLASSIFICATION: 435  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: CH 0016/97  
FILING DATE: 31-DEC-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Meigs, J. Timothy  
REGISTRATION NUMBER: 38,241  
REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976  
TELEPHONE: 919-541-8587  
TELEFAX: 919-541-8689  
INFORMATION FOR SEQ ID NO: 186:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 615 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:  
ORGANISM: PAG1074RP  
US-08-998-416-186

Query Match 0.6%; Score 39.4; DB 4; Length 615;  
Best Local Similarity 47.0%; Pred. No. 0.28;  
Matches 162; Conservative 0; Mismatches 176; Indels 7; Gaps 1;  
QY 5776 ATAAATCCATTATATGTAACAGCAATATTTAAGTTGAGACTTANGTGTAA 5835  
DB 185 ATAAATTAATTAATTAATCAATCAATTAATGCTATGTCATTAATTAATTAAT 244  
QY 5836 TTGATATATTTTGTGAGGTATACATATTTGGAATTCAGCTCAAAATGAGTCTCAG 5895  
DB 245 TAAATATTTATGATA-----TTATTTATTTCTTTAATAATTTATTAATGATTAT 297  
QY 5896 TATTAATTTAGATATCTTCATAGCAATGCTCTCTAAAGGTTTGTAAAGATATCAAT 5955  
DB 298 CAATTAATTAATTAATTTATTTATTAATGTTTAAATAATATATTTATTTATTAATAA 357  
QY 5956 GCCTTGATAGACTAATTTGTAGACTTTTATTTATTTCTTAAACCTTGATCT 6015  
DB 358 GATTAATTTATTAATAATGTAATTAATTTATTTATTTATTAATCTATTTTATTAATAA 417  
QY 6016 GCTATATAGTCATTATCTATCTATATAGATATGACGCCGTAGAGCAATCTCTGA 6075  
DB 418 TATTAATGCTATTTATTTATTTATTTAATCTTTTATTAAGAAATTTATTAATAATTTAA 477  
QY 6076 TTTTATATGTTTATTTCTTTCTTAATGAACCTTAGAAGACTA 6120  
DB 478 CTTTAATTTCTTATTTATTTATTTTATATTTATTAATAATTAATA 522

RESULT 36  
PCT-US95-06406A-11/c  
Sequence 11, Application PC/TUS9506406A  
GENERAL INFORMATION:  
APPLICANT: Janet D. Robishaw, Charles Kunsch  
TITLE OF INVENTION: CDNA Clones Encoding Human G Protein  
TITLE OF INVENTION: Subunits  
NUMBER OF SEQUENCES: 23

CORRESPONDENCE ADDRESS:  
ADDRESS:  
STREET:  
CITY:  
STATE:  
COUNTRY:  
ZIP:  
COMPUTER READABLE FORM:  
MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE  
COMPUTER: IBM 486  
OPERATING SYSTEM: WINDOWS FOR WORKGROUPS  
SOFTWARE: WORDPERFECT 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/06406A  
FILING DATE: Herewith  
CLASSIFICATION:  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME:  
REGISTRATION NUMBER:  
REFERENCE/DOCKET NUMBER:  
TELECOMMUNICATION INFORMATION:  
TELEPHONE:  
TELEFAX:  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 654  
TYPE: Nucleic Acid  
STRANDEDNESS: Single  
TOPOLOGY: Linear  
AMTI-SENSE: NO  
PCT-US95-06406A-11

Query Match 0.6%; Score 39.4; DB 5; Length 654;  
Best Local Similarity 56.6%; Pred. No. 0.29;  
Matches 73; Conservative 0; Mismatches 56; Indels 0; Gaps 0;  
QY 5242 TATTTTTCCTGTTTCTTTGTTGTTAAACCAACGTTACCTGATCATGCAATGA 5301  
DB 648 TTTTCTTTTCTTTTCTTTTCTTTTAAAGACAACTTATTTGAGCATCAAGCAATAT 589  
QY 5302 GAAGAAATTTTCATTTCAATTAATAGTCCCTTTGTTGATGACAGCTTGATC 5361  
DB 588 ACTGAAGAGTGTCATGCTCAAGTAATGTTCTTATTTCCATTAATTTGATCATGATACT 529  
QY 5362 AGCATCTCT 5370  
DB 528 CTCGCTCT 520

RESULT 37  
US-08-973-462-1  
Sequence 1, Application US/08973462B  
Patent No. 6191270  
GENERAL INFORMATION:  
APPLICANT: DRUILHE, PIERRE  
TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES  
FILE REFERENCE: 0660-0125-0 PCT  
CURRENT APPLICATION NUMBER: US/08/973,462B  
CURRENT FILING DATE: 1998-02-06  
EARLIER APPLICATION NUMBER: PCT/FR96/00894  
EARLIER FILING DATE: 1996-06-12  
EARLIER APPLICATION NUMBER: FR 95/07007  
NUMBER OF SEQ ID NOS: 29  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 1  
LENGTH: 6152  
TYPE: DNA



[illegible]

APPLICANT: Rogers IV, William O.  
 TITLE OF INVENTION: Protective malaria sporozoite surface protein  
 TITLE OF INVENTION: Immunogen and gene  
 NUMBER OF SEQUENCES: 2  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: A. David Spevack  
 STREET: NMHC Building 1 T-12 National Naval  
 STREET: Medical Center  
 CITY: Bethesda  
 STATE: MD  
 COUNTRY: USA  
 ZIP: 20814-5044  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent Release #1.24  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/07/638,431  
 FILING DATE: 19910110  
 CLASSIFICATION: 424  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Spevack, Avrom D.  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (301) 295-6759  
 TELEFAX: (301) 295-4033  
 INFORMATION FOR SEQ ID NO: 1:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 4673 base pairs  
 TYPE: NUCLEIC ACID  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 HYPOTHETICAL: N  
 ANTI-SENSE: N  
 ORIGINAL SOURCE:  
 ORGANISM: Plasmodium yoelii  
 STRAIN: 17X(NL)  
 DEVELOPMENTAL STAGE: erythrocytic stage  
 TISSUE TYPE: Blood  
 CELL TYPE: erythrocytic stage  
 IMMEDIATE SOURCE:  
 LIBRARY: Py-lambdaqtl1-2-7 kb genomic expression  
 CLONE: Py10.1111  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 718..3195  
 OTHER INFORMATION:

US-07-638-431-1

Query Match	0.6%	Score 38.2	DB 1	length 4673
Best Local Similarity	44.4%	Pred. No. 2.6		
Matches	195	Conservative	0	Mismatches 243; Indels 1; Gaps 1
QY	5676	TTTAATTGATGACGATATCAGCAGACATTTAACCTGCGCATGTTTTATTTTGT	5735	
Db	4195	TTAAATTAATTTATTTATTTTAAAGAACCAATATCATTTTCCAAATTTATTCATTTCT	4254	
QY	5736	TTTTTAGCAGT-GCTGACTAAGCCCAAGTTTTTGTAAAGATCAATAAATCCATTTATATG	5794	
Db	4255	TCATATTTGTAGTTGACATGCTCCATTTTGTGTTGGAGCATATAAATTTATTTATCT	4314	
QY	5795	TAAACACCAATTAATTTAAGTGAACAATTAAGTCTTTAATTTGATATAATTTTGGGAG	5854	
Db	4315	TTATTTGCTTAATCTTTATTTTGTAAATTTATTAATTTTGAAGATTTGATATATATTTCA	4374	
QY	5855	TATACATATTTGGAATTTGACTCAAAAATGAGAGTCTTCAGATTAATTTAGATATCTTC	5914	
Db	4375	TCCATTTTGTCTGCTTCATATTTATTAGAAAGAAATTAATTTCTATTTTATCAATA	4434	
QY	5915	ATAGCAATGTCCTCAAAAGGCTTTTGTAAAGATATCAATGCTTGATTAACCTAAT	5974	



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 12, 2002, 03:56:26 : Search time 867.35 Seconds

(without alignments)  
12185.767 Million cell updates/sec

Title: US-09-842-256-1

Perfect score: 6156

Sequence: 1 GCGGCGCCGACGCTCGGCTA.....CCACTTATGTTGTTT 6156

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : N\_Geneseq\_032802.\*

1: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.\*  
2: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.\*  
3: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.\*  
4: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.\*  
5: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.\*  
6: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.\*  
7: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.\*  
8: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.\*  
9: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.\*  
10: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.\*  
11: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.\*  
12: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.\*  
13: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.\*  
14: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.\*  
15: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.\*  
16: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.\*  
17: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.\*  
18: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.\*  
19: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.\*  
20: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.\*  
21: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.\*  
22: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.\*  
23: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.\*  
24: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6152.8	99.9	6156	19	AAV03517 Human transcriptio
2	597.6	9.7	4621	20	AAV99919 Murine PCIP gene (
3	547.4	8.9	4860	20	AAZ26000 Murine P/CIP gene.
4	542.8	8.8	6835	20	AAV99915 A1B1 (Amplified in
5	541.4	8.8	4789	20	AAH80992 Human steroid rece
6	510.6	8.3	6754	22	AAZ26480 Human SRC-3 DNA.
7	364.6	5.9	4660	21	AAA09325 Human cancer assoc
8	363.2	5.9	4547	22	AAZ03356 Human steroid rece
9	189	3.1	3806	22	AAK84490 Human immune/haema

10	187.4	3.0	3815	22	AAK84489 Human immune/haema
11	150.6	2.4	767	22	AAH04011 Human cDNA clone (
12	150.6	2.4	1367	22	AAH7184 Human cDNA sequenc
13	106	1.7	435	22	AAH9288 Human protein encoc
14	95.4	1.5	3361	20	AAH80993 Human breast cell
15	70.6	1.1	330	22	ABA68512 Human foetal liver
16	70.6	1.1	330	22	ABA35492 Probe #13958 for g
17	70.6	1.1	330	22	AAK16879 Human brain expres
18	70.6	1.1	330	22	AAK42647 Human bone marrow
19	70.6	1.1	330	22	AAI23403 Probe #13336 for g
20	70.6	1.1	330	22	AAI48724 Probe #17410 used
21	70.6	1.1	330	22	AAI09030 Probe #9021 used t
22	64.8	1.1	3388	18	AAH84543 Mouse neuronal PAS
23	58.6	1.0	2082	19	AAV41257 Human immune syste
24	58.6	0.9	6131	24	ABL22890 Tumour suppressor
25	56.8	0.9	7135	22	AAH6423 Human immune syste
26	56.4	0.9	8946	24	ABL32911 Human gene regulat
27	55.6	0.9	6418	24	ABL32322 Human neuronal PAS
28	55.6	0.9	6418	24	AAH61073 Human clock gene.
29	55.4	0.9	1581	20	AAH8984 Human HSCLOCK cDNA
30	55.4	0.9	2078	19	AAV41256 Human transcriptio
31	55	0.9	3486	21	AAZ34629 Human cancer associ
32	55	0.9	3546	19	AAV61450 Chemically pretrea
33	55	0.9	5715	20	AAH03456 Human immune syste
34	55	0.9	6161	24	ABL32623 Human immune syste
35	54.2	0.9	4260	21	AAH18177 Human transcriptio
36	53.8	0.9	6665	22	AAH45298 Lung cancer associ
37	53.8	0.9	6665	22	ABL32082 Human immune syste
38	53.6	0.9	15373	24	ABL32466 Human immune syste
39	53.6	0.9	6175	24	ABL33307 Human immune syste
40	53.4	0.9	3614	18	AAH91883 Murine SIM (single
41	53	0.9	6065	24	ABL32504 Human immune syste
42	53	0.9	18154	24	ABL2254 Human immune syste
43	53	0.9	19087	24	ABL2793 Human immune syste
44	53	0.9	19087	24	ABL2793 Tumour suppressor
45	52.8	0.9	5947	22	AAH4676

# ALIGNMENTS

RESULT 1	AAV03517	standard; cDNA: 6156 BP.
ID	AAV03517	
XX	AAV03517	
AC	20-JUL-1998	(first entry)
DT		
XX		
DE	Human transcriptional intermediary factor-2 (TIF2) cDNA.	
XX		
KW	Transcriptional intermediary factor; TIF2; human; drug screening;	
KW	assay; nuclear receptor; ds.	
XX		
OS	Homo sapiens.	
XX		
FH	key	Location/Qualifiers
FT	CDS	163..457
FT		/*tag= a
XX		
PN	WO9802455-A2.	
XX		
PD	22-JAN-1998.	
XX		
XX	11-JUL-1997;	97WO-US12100.
PF		
PR	12-JUL-1996;	96US-0021247.
XX		
PA	(BRIM ) BRISTOL-MYERS SQUIBB CO.	
PA	(CNRS ) CENT NAT RECH SCI.	
PA	(INRM ) INST NAT SANTE & RECH MEDICALE.	
PA	(VIPA-) UNIV PASTEUR LOUIS.	
XX		

PI Chabon P, Gronemeyer H, Lutz Y, Voegel J:  
XX WPI; 1998-110525/10.  
DR P-PSDB; AAW42632.  
XX  
PT New isolated transcriptional intermediary factor-2 - useful in  
PT assays for drugs which are capable of enhancing or inhibiting  
PT nuclear receptor-mediated pathways  
XX  
PS  
XX  
XX Claim 1; Fig 1A-B; 119pp; English.

This cDNA clone codes for a novel nuclear receptor transcriptional  
CC mediator (see AAW42632) designated transcriptional intermediary  
CC factor-2 (TIF2). TIF2 interacts directly with the ligand binding  
CC domains (LBD) of several nuclear receptors (NR) in an agonist- and  
CC AF2-integrity-dependent manner in vitro and in vivo, harbours an  
CC autonomous AF, relieves NR autoquenching, and enhances the  
CC activity of NR AFs when overexpressed in mammalian cells.  
CC Screening of a human placenta cDNA expression library with an  
CC estradiol-bound 32P-labelled estrogen receptor probe yielded a cDNA  
CC that encoded a protein fragment (TIF2.1) that interacted on  
CC Far-Western blots with different 32P-labelled NR-LBD ER, RAR, RXR  
CC in an agonist-dependent manner. The TIF2 coding sequence was  
CC obtained upon rescreening with a TIF2.1 cDNA probe. TIF2 cDNA has  
CC been deposited at ATCC 97612. Polynucleotides encoding full-length  
CC TIF2 protein or encoding TIF2 polypeptide fragments comprising  
CC amino acids 624-869, 624-1131, 1010-1131 or 1288-1464 are claimed.  
CC Recombinant methods for making TIF2 polypeptides are provided, as  
CC are screening methods for identifying agonists and antagonists of  
CC nuclear receptor AF-2 function, TIF2 AD1 activity and TIF2 AD2  
CC activity. The products are useful in assays for identifying drugs  
CC capable of enhancing or inhibiting NR-mediated pathways. They can  
CC also be used for detection and localisation.  
XX  
XX

Sequence 6156 BP; 1799 A; 1490 C; 1406 G; 1461 T; 0 other;

Query Match 99.9%; Score 6152.8; DB 19; Length 6156;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 6154; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GGGGGCCGACGCTCGGCTACAGTTCGGGGGGAAGGTAGCGCCGACCGCGCA 60  
DB 1 gggggccgacgctcggtctacagttcggggggaaggtagcgccgacccgca 60  
QY 61 CCGAGCGCGGTGACGACCGCGGACGATTCCTGGATTGGCTACACACTATAGATC 120  
DB 61 ccgagcgcggtgacgacccggaacgattctcttgattggtctacacctatagatc 120  
QY 121 TTCTGCACTGTTTACAGGACAGTTCGTATGTGTTCAAGATGAGTGGATGGAGAA 180  
DB 121 ttctgcactgtttacaggacagtctgtatgtgttcaagatgagtggtggatggagaa 180  
QY 181 AATACCTGACGCCCTCCAGGCGAGACAAAGAGCGAAGTCTCTGACCAACTT 240  
DB 181 aatacctgacgccctccaggcgagacaagaagcgaaatgctctgaccaactt 240  
QY 241 GGAGCCAGGCCCAAAAGAACTGAATAAAGCTAATCGTGAACAGGAAATAATATA 300  
DB 241 ggagccaggcccaaaagaaactgaaataaagctaatcgtaaacaggaataatata 300  
QY 301 GAAGAACTTGACAGATTGATTTTGAATAATTAATGATATAGACAATTAACTTCAA 360  
DB 301 gaagaaacttgacagattgattttgaaataatgaatgatatagacaacttaactcaa 360  
QY 361 CCGAGCAAAATGTGCAATCTTAAAGAAACGTGAAGCAATGCTGACATCAAGAACAA 420  
DB 361 ccgagcaaaatgtgcaatctttaaagaaacgtgaaagcaaatgctgacatcaagaa 420  
QY 421 GAGAAAGCAGCAGCTGCCAATAGATGAAGTGCAGAACTGATGATCTTACAGG 480  
DB 421 gagaagcagcagctgccaatagatgaagtgcagaagtcagatgtatctcttacaagg 480

QY 481 CAGGCTGTCATGCACAAGATGCGGCTGGGGCTATGATGCTTAGGCCCTTATGSGTTC 540  
DB 481 caggctgtcatgcacaagatgcggctggggctatgatgcttagggccttattgsgttc 540  
QY 541 TTCTTGTAAGTGAACCTGGAAGGCAACGTGTCTTGTGACAGAAATGACAGTAT 600  
DB 541 ttcttgtaagtgaacctgggaaggcaacgtgtcttgtgacagaaatgacagtat 600  
QY 601 CTAAAGTATTAACCAAGAGAGCTGATGAACAAAGTATATATGACATCTTGATGGG 660  
DB 601 ctaagtatataccaagaagagctgatgaacaagaatgatatgacatcttgatggg 660  
QY 661 GACCAACAGGAATTTGTCAAAAACCTGCTCCAAAGCTATAGTAATGGGGATCTGG 720  
DB 661 gaccacaagaaattgtcaaaaaacctgctccaaagctataagtaaatggggaattcgg 720  
QY 721 TCTGGGGAACCTCCGAGGGCGGAACGCAATACCTTCAATGTGCGAGTGTAAACCT 780  
DB 721 tctggggaacctccgagggcggaacctcaataccttcaatgtcgagtggttaacct 780  
QY 781 TTACTGATTCAAGAGAGAGGCTCATGATTAACAGAACTCATGAAATATGAACAT 840  
DB 781 ttactgattcaagagagagggtcatgataaccaggaagctcatgaataatgaacct 840  
QY 841 ATCCAGTCTTCCCTGCTCTCTCAACCAAGTCCATCAAGAGAGAGAGATTTGACG 900  
DB 841 atccagtcttccctgctctctcaacaaagtcacataaagaagaagaatcttgacg 900  
QY 901 TCTGCTGATGTTTGGCGGCAAGAGATTCCTATGAGGAAGACAGATTCCTCCCTCA 960  
DB 901 tctgctgatgtttggcggcaagagatctctatgaggaagacagatctctccctca 960  
QY 961 TCAGAAAGTTTACTACTCCAGATCTCCAGATCTCCAGAGCAAGATCAGTCTGATAC 1020  
DB 961 tcagaaagtttactactccagatctccagatctccagagcaagatcagctctgataccg 1020  
QY 1021 ACCATGAGAGCACCATGAACCAAGCGCTGGAGACCTGGTAAGAAAGTGTATGAAAG 1080  
DB 1021 accatgagagcaccatgaaccagcgctggagacctggttaagaaagtgatattcagaag 1080  
QY 1081 TTCCATGCGCAGCATGAAGAGATCTGTCTATGCTATGAGGAGATCATGATGAAGTA 1140  
DB 1081 ttccatgcgccagcatgaagagatctgtctatgctatgagggagatcatgatagaagta 1140  
QY 1141 CTGAGCAAGAGATTGGCATTCATCAAAATCTATGCTTTTCTGTCGTGACACTT 1200  
DB 1141 ctgagcaagagattggcattcatcaaaatctatgcttttctgctgtgacactt 1200  
QY 1201 GTTGCTGACAAAGAGAGCAAACTCATCTCGTTTCAGCTACTAATGAACTCAACT 1260  
DB 1201 gttgctgacaaagagagcaaaactcatctcgttctcagactactaatagaacctcaactt 1260  
QY 1261 GTAATATCTTATCATATGCTTACAGAGACAGCAAAATGCTGTGATGAAATCCGATCTG 1320  
DB 1261 gtaatatcttatacatatgcttaccagagagcaaaatgctgtgataaaccgagctcg 1320  
QY 1321 ACTGACAAAAGATGGGGAAGCCACTGAATCAATTAAGCTTAACAGCCCTGCCATCA 1380  
DB 1321 actgacaaaagatggggaaagccactgaatcaattaaagcttaacagccctgccatcag 1380  
QY 1381 GCCCTGTGAGTGGGGAACCCAGGTCAAGACATACCTCTAGTAGCAATATAAATTTCC 1440  
DB 1381 gccctgtgagtggggaacccaggtcaagacatacctctagtagcaaatataaatttccc 1440  
QY 1441 ATAAATGGGCCCAAGGAACAAATGGGATGCCATGCGGAGTGTGGTGTGGGGG 1500  
DB 1441 ataaatgggcccaaggaaacaaatgggatacgccatgcggtgtgggtgtggggg 1500  
QY 1501 ATGAACCATGTGTCAAGGATGACGAACCACTCTCAGGGTATGATGACATCAAA 1560  
DB 1501 atgaaccatgtgtcaggatgacgaaccaactctcagggatgataatgacatcgaacc 1560  
QY 1561 ATGAACAGCCCTCTCAACAAAGCAGCCCTGGCATATTCAGAGACGCCACCTCATGCTT 1620



Db 1561 atgacagccctcacaagaagccctgcatgataccagacagcccaactcattgctt 1620  
Qy 1621 TCACCAAGGCAATCGCATATGACCCCTGGAGTGGCTGGACACCTCGAATCCCAACCACTAG 1680  
Db 1621 tcaccaaggaatcgcatgagccctgagtgctgagcagccctcgatcccccagctag 1680  
Qy 1681 TTTTCCCTGAGAGAGTTTCATTTCCCTGGAGGAGTTTCAGAGCAGCAGGAATAGC 1740  
Db 1681 ttccccccgaggaagctcgtcattccccctgaggagcttcgacagacaggaataagc 1740  
Qy 1741 CATAGTATACCAACAGCTCCCTCATGTCATTCAGGCCCTCAGCGAGGGCAGCGGGTC 1800  
Db 1741 catagtataacacaagctccctcctaagctcactcaagccctcagggagggcaggggtt 1800  
Qy 1801 TCATTAGGCTATGCTTGTTGGCTTCACAGACTAAAAATGGCAATTTGCCAAAATCTCCCA 1860  
Db 1801 tcattagggatcgttgccttcacagacctaaataatgggcaatttcgcaaaactcccca 1860  
Qy 1861 GTTAAATGAAATCCTCCCTCAGCAAGATGGAGCTTGAGCTCAAAAGACTGTTT 1920  
Db 1861 gttaatatgaatccctcccccactcagcaagatgggaagcttgcacaaagactgctt 1920  
Qy 1921 GGACTATATGGGAGCCCTCTGAAGGTACAACTGGACAGCAGAGCAGCTGCCATCT 1980  
Db 1921 ggaactatattggggagccctcgtgaaggtatacaactcgacacagagagagagctgcact 1980  
Qy 1981 GGAGACGAAAAGAAACAATGACCCCACTGCCCGCCGCTGAGCAGTGGAGAGCT 2040  
Db 1981 ggaagagaaaaagaaacaaatgaccccccaactcgccccgctgtagagagtgagagct 2040  
Qy 2041 GACGGGCGAGCAGACACTCATGACAGCAAGAGGCGAGACCAACTCCTGACTGCTGAC 2100  
Db 2041 gacgggagcagacagactgcgatacagcaagagggcagaccaaactccctgcagctctgccc 2100  
Qy 2101 ACCAATCTGATCAGATGAGAGCCCTCGCCCTTAGCCAGACTTTTTCGGATACAAACAA 2160  
Db 2101 accaaatctgactcagatgagccctcgcccttagccagctcttctgcatatacaaaaa 2160  
Qy 2161 GACGCCACAGTACTGCTGCCGTGTTCTGGGTCTACACTGGAACCTCGCTCAAGAGAG 2220  
Db 2161 gacgcccaagtgactgctgctggtctcgtggtctacaatgaaacctgcctcaagagagag 2220  
Qy 2221 CATAAATTTTGGCAGACACTTTCAGAGCAGCAGATTCCCTGTGGATTGGCAAGTTA 2280  
Db 2221 cataaaatttgcacagactctgcagagacagcttccctgtagacttgcagcaagta 2280  
Qy 2281 ACAGCAGAAAGCCACAGGCAAGACCTGAGCGAGAGTCCAGCAGCAGCAGCTCTGGATCA 2340  
Db 2281 acagcagaagccacagagcaaaagacctgagccagagctcagcagcagcctccgtgatca 2340  
Qy 2341 GAAATGACTATTAAACAAGACCCGCTGAGCCCAAGAAAGAAAGATGCACTACTTGGC 2400  
Db 2341 gaagtgcatactaaacagagccggtgagccccaagaaagaaatgacactacttctgt 2400  
Qy 2401 TATTGCTAGATAAGATGATACTAAAGATATTGTTTACAGAAATTAACCCCAACTT 2460  
Db 2401 tatttgcagtaagaatgataactaaagatatgtgttaaccaagaataatacccccaactt 2460  
Qy 2461 GAGAGACTGACAGCTAAGACAGATCTGCGAGTAACACAAAATTAAATGCAATGAAAAT 2520  
Db 2461 gagagactgagcagtaagaacagatccctgcagtaacacaaataatagcaatgaaact 2520  
Qy 2521 GAGAAAGAGGAGATGAGCTTTGAGCCTTGATGACCAAGCTTGCGAGTGGAGCTGGACAACTTG 2580  
Db 2521 gagaagagagagatgagccttgcagccctgggtgacagccctgcagtgagctggaacacttg 2580  
Qy 2581 GAGCAGATTTTGGATGATTTGCAGAAATAGTCAATTACCAAGCTTTTCCAGACGAGG 2640  
Db 2581 gagagagatttgcagatcttgcagaaatagtaacttaaccaagcttccccaagacaggg 2640  
Qy 2641 CGAGGCGCCCTGCTGATCAGTTGACAAGCAAGCATCATCAATGACCTCATGCACTC 2700  
Db 2641 ccaagcgccctgctgatactgatacgaagcaagccatcatcaatgacatcatgcaactc 2700  
Qy 2701 ACAGTGAAGAAACAGCCCTGTCACACACTTGTGGAGGCCCAAGAAACAGCACTCCGAATTTC 2760  
Db 2701 acagctgaagaaacagccctgtacacacctgttgagagcccaagaaacagacgctgcaatttca 2760  
Qy 2761 CAGACACTTTTAAATTAACCCAGACCAAGGCAACTGGGCGAGGTTATTTCCAAACAGAT 2820  
Db 2761 cagagcaactttaataaacccacagcagaggaacttggcaggttatctgcacaaacagaaat 2820  
Qy 2821 TTACCACTTGACATCACATTGCAAGCCCAACTGGTGGCTGAGCACTTTCCACCAATCAGA 2880  
Db 2821 ttaccacttgcatacaacttgcacaaagcccaactgtgtgttgcaccttccccaactcaaga 2880  
Qy 2881 AACAGTACTCCCTACTCATGATATACCTAGCCAGCAATGATGGGTAAATCAAGGATGATA 2940  
Db 2881 aacagtaagcttccctactcagtaataccttaagcaggaatgaaatgggttaatacaaggatgata 2940  
Qy 2941 GGAACCAAGGAAATTTTGGGAGACAGTAGCAGCAAGATGATTGGTAAACAGTGCCTTCGG 3000  
Db 2941 ggaacccaaggaatcttagggaaacagtagcaaggaatgattggtlaaacagtgcttcgg 3000  
Qy 3001 CCTACTATGCACTGGAAGATGGGCAACCGCAGAGATTCGAGCTGTGAGAGTCACTGTGCT 3060  
Db 3001 cctactatgcactctgaggaattgggcaaccgaggttcgctgtgagagtgacactgtgct 3060  
Qy 3061 GCTACACCAAGTGCATATACCGGCGCACTCCAGAGGATGATTCGGAACCCAGCAGCC 3120  
Db 3061 gctacccaagtgccatgaaacccgagccagctccaagaggtatgatttcggaacccagcagcc 3120  
Qy 3121 AGCATCCCATGAGGCGCCAGCAGCAGCTGGCCCAAAACAGAGCCTTCAGTCTCAGGTC 3180  
Db 3121 agcatccccaatgagagcccaagcagcttggccaagaacagagcttcagttccaggtc 3180  
Qy 3181 ATGAATATAGGGCATCTGAATTAGAGATGACATGGGGGACCTCACTATAGCCAACAA 3240  
Db 3181 atgaataatagggcatctgaaattagagaatgaaatgagggagccctcaglatagcaaaaa 3240  
Qy 3241 CAAGCTCCTCAAAATCAGACACTGCCCAATGGCCTGAAGAACCTCGCCATAGCAGGCG 3300  
Db 3241 caagctccccaatcaagactgcgcccagctgagaaatcctgcctatagacagagcg 3300  
Qy 3301 TCTTTTGGCAGCCAAACAGGCGAGCAGCAATTTGGAGTTTCCAGATGACTTCTATGTTCA 3360  
Db 3301 tcttttgcagcccaaaaagcagagcatttgcaggttccagatgacttgcatagttca 3360  
Qy 3361 CATCTCGAGCTGAGTCTCCAGTGATGAGGAGCTCTCTGGACAGCAGCTATCTGSCC 3420  
Db 3361 catctcgagctgagctcgcagatgagtgagagctctccctgcagcagctgtatctgccc 3420  
Qy 3421 TTGCGGAATTTTGAATGGCTGGAGAGATTGATAGACCTTAGAATATCCGGAATGGTC 3480  
Db 3421 ttgcggaattttgaatggcctggagagatttgatagaccttagaataatccgaactgtc 3480  
Qy 3481 AGCCAGAGCCAAAGAGTGTGATCTCAGAAACAGTTTCAATTCAGATTCGAACATATGCTG 3540  
Db 3481 agccaagcccaagcagtagatccagaacagtttccaagtcaggttccaatacatatgctg 3540  
Qy 3541 GAGCAGAAAGCGCCGCTTTTCCACAGCAGTATGCATCTCAGGCAAAATGGCCCAAGGT 3600  
Db 3541 gagcagaagcgccgcttttcccaagcagatgcatctcaggaaccaaatggcccaaggt 3600  
Qy 3601 AGCTATTCTCCATGCAAGATCCAAACTTTTCAACACATGGGAACAGCGGCTTAAGTTATGCC 3660  
Db 3601 agctattctccatgcaagatcccaacttcaacaccaaggagcagcggtcctagttatagcc 3660  
Qy 3661 ACACCTCCGTATGACGCCAGACCGGGCTCAGGGCCACAGGGCTTATGTCAGAAACAGCA 3720  
Db 3661 acactccgtatagcagcccaagccggctcagggccccaagggccttagtgcaaaacagca 3720  
Qy 3721 AATCAACTAAGACTTCAACTTCAGCATGCTTCCAGACAGCAGAAATGCGCAGCACCTT 3780  
Db 3721 aatcaactaagaactcaactcagatcgctccaagaagacagcagaatcgcagcagcaact 3780

OY 3781 ATGAATCAATCAGCAATGTTTCCAAATGTGAATCTGACTGTGAGGCGCTGGAGTACCAACA 3840  
|||||  
Db 3781 atgaatcaaatcagcaatgtttccaaatgtgaatctgactgtgagcgctggagtaaccaaca 3840  
OY 3841 CAGGCACTATTAAATGCACAGATGTGGCCCGACAGACAGAGGGAATTCCTGAACAGCAT 3900  
|||||  
Db 3841 caggcaccattaatgcacagatgtgcccagagacagaggaatccctgaaccagcat 3900  
OY 3901 CTTTCGACAGACAAATGATCATCGCAACGCAATTGACGAAGCACTTGTGATGATGAG 3960  
|||||  
Db 3901 ctctgcagagacaatgcatcgcagcaacagcaagttcagcaacgaacttgatgatyaga 3960  
OY 3961 GGACAGGGTTGAATATGACACCAACGATGTGGCTCTAGTGTATGCGACGACACTATG 4020  
|||||  
Db 3961 ggacaaagggttgatataatgacacaaagcatgtgtcctcagtgatgtgcagaactatg 4020  
OY 4021 AGCAACCTTCGGATTCCTCCACGCAATGCAACAGACTTTCCTCCAAACTACGGA 4080  
|||||  
Db 4021 agcaaccttcggatctcccgagcaaatgcaagcagtttcacttctccaaactacgga 4080  
OY 4081 ATTAAGTCAGCAACCTGATCAGGCTTACTGGGGCTTAGCACTGCCAGAGCCCACTTATG 4140  
|||||  
Db 4081 ataagtcagcaacctgatalccaagcttctacggtgagcaagctcccaagcccaactatg 4140  
OY 4141 TCACCCCGAATGGCACATACACAGAGTCCCATGATGCAACAGTCTCAGGCAACCCAGCC 4200  
|||||  
Db 4141 tcaccccgaaatggcacatcacagagttccatgtgcaaacgctcgaagcccaaccagcc 4200  
OY 4201 TATCAGGCCCCCTCCGACATAAATGATGGCGCAGGGGAACATGGCGGAAACACACTG 4260  
|||||  
Db 4201 tatcagggccccctccgacataaattgagctggcgagggaacataggcggaacaaacagcatg 4260  
OY 4261 TTTTCCAGAGATGCCACACACTTTGGGACAGCAAAACACGACATGTACATTAAC 4320  
|||||  
Db 4261 ttttccagagatgccacacactttgggacagcaaaacacgacatgtacatgtaaac 4320  
OY 4321 AACATGAACATCAATGTGTCCATGGCGACCAACAGGTGGATGAGCAGCATGAACAG 4380  
|||||  
Db 4321 aacatgaacatcaatgtgtccatggcgacccaacaggtggatgagcagcatgaaacag 4380  
OY 4381 ATGACAGGACAGATGACGATGACCTCAGTGACCTCGGTGTCTACGTACAGGGCTGTCTCC 4440  
|||||  
Db 4381 atgacagggacagatgacgactgacctaagtgacctcgtgtctacgtacagggctgtctcc 4440  
OY 4441 ATGGATCCGAGCAGGATTAAATGATCTGTGCTGTGAGGGAAGCAACTGTCCCAACAG 4500  
|||||  
Db 4441 atggatcccgagcaggttaatgattcgtctgtcgtgaggggagcaacctgttcccaaacag 4500  
OY 4501 CTGCTGGAATGGATATGATTAAGACAGAGGAGACAAACACGAAATATTGCTGACAC 4560  
|||||  
Db 4501 ctgctggaaatggatatgattaaagacagagagacaaacgaaatattgctgacac 4560  
OY 4561 TGCTGAAGGCAGTTGCTTCTTCAGCTGACGAGGCTCATCTGCTCAAAACCTTCCAGTCT 4620  
|||||  
Db 4561 tgctgaaggcagttgcttcttcagctgacgggctcactgtctcaaaaaccttccagttct 4620  
OY 4621 GGAGACTGTGCTATTGTTGTTCAACCAACTACCTGCGACGCGGTTGCTGTAGAGCAG 4680  
|||||  
Db 4621 ggagactgtgctatttggtttcaaccactacactgacgacggggttctgtcagaagcag 4680  
OY 4681 ACAGGCTTGCCCTGTGTTCCAGAGGTGGCGTCCACTCGGCTGTGGCAGAGAGAGTGCTT 4740  
|||||  
Db 4681 acaggcttgccctgtgttccagaggtggcggtccactcgtgtggcagggagcgctgtgctt 4740  
OY 4741 CTTCCTCTGACACTGGAAGCTTGCAATCCAGACAGTCCGCTCACTGTTCCTGCATTTCA 4800  
|||||  
Db 4741 ctctctctgacactggaagcttgcaatccagacagctgcactgttcttcccgcatlca 4800  
OY 4801 CCTTAGTGAACCTTAGATCTCTCTCTCCCAAGTAATGTTGACAGGCAATTTTCATACCC 4860  
|||||  
Db 4801 ccttagtgcaaccttagatctctctctcccaagtaaatgttgacaggccaatttcatacc 4860

OY 4861 ATGTCAGATTGAATGATTATTTAAATGTATGTATTTAAAGAGAAACCATGCTTGTCTGT 4920  
|||||  
Db 4861 atgtcagattggaatgattatftttaaattgtatgtatttaaggaacacatgcttctgttct 4920  
OY 4921 CCTGTGCGGTTCCAGACACTGGTTTCTTGCTTTGTTTCCCTGGCTTAACAGTCTAGTGC 4980  
|||||  
Db 4921 cctgttcggttccagacactggtttcttgcctttgttlttccctggcttaacagctagtcg 4980  
OY 4981 AAAGATTAGATTATTCGTGGGGAGAAAGAAATTTTAAAAATTAATAAATAAGAT 5040  
|||||  
Db 4981 aaagattagattatcttcgtggggagaaagaaatftttaaataataactaaatagat 5040  
OY 5041 GTTTTAAGCTAAAGCCTGATTTGGGATGGAAGCAGACAGACACCGTGGACAGCGCTGT 5100  
|||||  
Db 5041 gtttaagctaaagcctgatttgggaatgggaagcagacagaccgttgacagcgctgt 5100  
OY 5101 ATTTACAGACACACCAGTGGCGGTGAAGACCACAAAGTACAGTGTATCTCTAGAAAC 5160  
|||||  
Db 5101 atttacagacacaccagtgctggaagaccacaaagtcacagctgtacttctctagaagc 5160  
OY 5161 TCTAAGACCATGTTGGAAGAAGTCTCAGTTACTGAACAGATGAAAGAAGACCTGTGAG 5220  
|||||  
Db 5161 tctaagaccatgattggaagaagctccagttactgtaacagatgaaagagcgctgtgag 5220  
OY 5221 AGGCTGTTAACAATTAGCAAAATATTTTTCCTGTTTTCCTTTGTTAAACCAACTGG 5280  
|||||  
Db 5221 aggctgtttaacaattagcaaaatatttccctgttlttcttctgtttaaacaactgg 5280  
OY 5281 TTTACCTGTAATCATGAAATTTGAGAGAAATATTTTCAATTTTCAATTAAGTCCCTTTAG 5340  
|||||  
Db 5281 ttacactgtaatcatgaaatfttggagaaatatttcaatfttcaatlttaagtcoccttgg 5340  
OY 5341 TTTGATCACAACAGCTGAATTCAGCATCTCTTCTCCCTGTCAGCCGTCTCTCCCTTTC 5400  
|||||  
Db 5341 ttgatacagacagctgtgaatcagcatctcttccctgttcagcctgacatcttcccttc 5400  
OY 5401 CCTCTCTATTCCTCCATCTCCCTATTTTCAATTCCTTTTAAATAATATTAAGCTAC 5460  
|||||  
Db 5401 cctctctatctcccatctacatccctatttccatcttcttlttaaaaaataataagctac 5460  
OY 5461 AGAAACAGATTAAGCCCTTATTTCTTAAATGTTTGGCAGCAATTTACCAATGCTAA 5520  
|||||  
Db 5461 agaaacagattaaagcccttatttcccttaaatgttcttcagccacttaccatctgctaa 5520  
OY 5521 GTATTGAATTTTCAAGAAAAAAATGCAATTTACTGGCAAGGAGAAAGCAAAATTAAAGC 5580  
|||||  
Db 5521 gtaattgaattttcagaaaaaaatgcattttacttactgccaaggaaggaagtttaagc 5580  
OY 5581 TTGATACCAATGAGCTTAAGGATACCTGCTTTGGAAGCATGTTTATTCTGTTCCAGCA 5640  
|||||  
Db 5581 ttgataccatcgagctlaaggaatacctgtcttggaaagcatgttacttctgttcccaaga 5640  
OY 5641 ACTCTGCGCTCCAAATTTGGAGAAAGCCAGTGTGTTAAATTTGATAGCAGATATACGA 5700  
|||||  
Db 5641 acctctgcccctcaaatgggagaaacgcagtgctttaaattgaagagatataca 5700  
OY 5701 CAGATTTAACCTCTGCAATGTGTTTATTTTATTTGTTTAAAGCAGTGCAGCTAAAGCCGA 5760  
|||||  
Db 5701 cagatttaacctctgcacatgtgttcttatttgttctttagaagtgctgcactaagcga 5760  
OY 5761 AGTTTGTAAAGTACATAAATCCAAATTTATATGTAAACAGCAATATTTAAAGTTGACA 5820  
|||||  
Db 5761 agtttgtaaagtacataaattccaattatataatgaaagaacaaatlaagtgtgaga 5820  
OY 5821 ACTTATGTGTTTAAATTTATATTTTGTGAGGTATACATATTGGAATGACATCAAA 5880  
|||||  
Db 5821 acttattgtttaaattttaaatttcttgtaggtatacatatctggaattgaaaccaaa 5880  
OY 5881 AATGAGTACTTCAATTAATTAATATCTTCATAGCAATGTCTCCAAAGAGTGTATT 5940  
|||||  
Db 5881 aatgagttacttcaatataatataatcttcatagcaatgtctccctcaaggtgttct 5940  
OY 5941 GTAAAGATATCAATGCTTGATTTAGACCTTAATTTGTAGACTTTTATTTTCT 6000



[illegible]

Db	2044	caccagccctctcggagtgctccctcccaacatcaggaagtgcttccactccaatgct	2103
Oy	2196	ACATGGAACCTCGCTCAAGGAGAAACATATAAATTGTCACAGACTTTGCAGACGACGAC	2255
Db	2104	gcatgggtctctgtctcgaagaagaacacggatlttgcacaagttgcgtcgcgaatgycgaa	2163
Oy	2256	TTGCCCTGTGGACCTTGGCCAAAGTTAAACAGACGAGAAGCCACAGGCAAAAGACCTGAGCCAGCA	2315
Db	2164	ctcccacgaggagtgctgcagaagatcactgcagagggccacttggga-----agga	2211
Oy	2316	GTCCACGACGACACAGCTCTTGATCAGAAGTGACTATTTAAACAAGACCGGTGAGCCCCA	2375
Db	2212	cacgagcagactgtctctcgtctgagagggacaacacggcgcagcgactgagctctaa	2271
Oy	2376	GAAATAAGAG---AATGCCACTACTTGGCTATTGCTAGATAAAGATGACTAAGATAT	2432
Db	2272	gaagaaggaataataatgctctgtctagtataccgtgcgacggagatgacccagtgatgt	2331
Oy	2433	TGGTTTACCGAATAATACCCCAACTTGAGACACTGCAGACTAA---GACAGATCTCTC	2489
Db	2332	gtctgcacaagaagctctgagccccaagccgaacgctgggagcagtaactgagtcagtgca	2391
Oy	2490	CAGTAAACACAAATTAATTAGCAATGAAAACTGAAGAGGAGATGAGCTTTGACCTCG	2549
Db	2392	ctgtcccaaccaatcccgagctctgcgaagaagaccccaaatlaagaaccggaacgaa	2451
Oy	2550	TGACCACGCTGGCAGTGAGCTGGCAACTTGGAGGAGATTTTGGATTTGACATATAG	2609
Db	2452	cgaaggaagatctcgggaacctggaactaatactagatgcactcttgagatcttgacaagttc	2511
Oy	2610	TCAATTACACAGCTTTTCCACAGACACAGAGGCGCGCCCTGCTGATCAGTTGACAA	2665
Db	2512	tgacct-ctcaacaactcttaacaatbvgcgtcacccaggggccaacacagcagatgtctg	2570
Oy	2670	GCAAGCATATCAATATACCTCATGCACACTCAGCTGAAAAACGCCCTGTACACCTGT	2728
Db	2571	caggacggaagttcttcvgggtctgcgaagttccag-----ccgtgacgtctgt	2619
Oy	2730	TGGAGCCCGAGAAAACACACACTGGGAATTTCCAGACACACTTTTAATAACCCAGCACAG	2785
Db	2620	tcgtctccataataaccgagcggtgtctctgtagccctgtctgtgtgctcagttcc	2679
Oy	2790	GCAACTGGGCAAGGTTATTGGCAAAACCAAGATTATACCACTTGACATCACATTGCCAAAGCC	2849
Db	2660	gcacgtgaagaatgctcagtgcttccctcgtgtctacaaacacgcccatactggtctggaa	2739
Oy	2850	AACGTGACTGAGACTTTCCACCAATCACAAAACAGTACTCCCTACTCTGATGATCTCA	2905
Db	2740	tccaagaatggtgtagtagtcagaagaattacggtgcacaacatbvggcccacaacagaatgct	2799
Oy	2910	GCCAGGAATGATGGGTATATCAAGGGATGATAGAAACCAAGAAATTTAGGACACATAG	2966
Db	2800	tcctgtgaatccgacttctcccccggagactcgtggacttagacttaactcaaggccagcag	2855
Oy	2970	CACAGGAATGATTTGTAACAGTGCTTCTCG-----GCCATTATGCCATCTGGAG	3019
Db	2860	aatggaagctcttgcatcaaggtccctcgtggaaagaaactggaagccgatatacagtgccaatt	2919
Oy	3020	AATGGGACCCGACAGTTGGCTGTGAGACTCACTGTGCTGTACACACAGTGCCTATGA	3075
Db	2920	accagagacttgccaatggggggtctctgaccttaccttgcaacttcgtcttaactcagctgc	2979
Oy	3080	ACCGGCAAGTCCAAAGAGGATATGATTGGAACCCAGACACGCAAGCAATCCCATGAGGCCA	3133
Db	2980	aggtgtcaagaccatcgtlttgacaagaacagcagcaacagcagcaacgcaacaaca	3035
Oy	3140	GCAGC-----CAGCTGCGCCAAAGACAGACGCTTCAAGTCTCAAGTCAATATATAGGCGC	3194
Db	3040	gcagacaacagcagcaacaacagcagcagcgaacagcagcagatgtcttcaatgagaac	3099
Oy	3195	ATTCGATTTAGAGATGAACATGGGGGACCTCAGTATAGCCAAACAAGCTCTTCAAA	3254

D	b	3100	tggtagattcccatggygaatggygagtcaatccctatagc-----caagcagtgccgtctaa	3150
O	y	3255	TCAGACTGCCCATGGCCTGAAAGCATCCGCTATAGACAGGCGCTTTTGGCCAGCCA	3314
D	b	3157	ccaacacgattctccgacgaaggcagctctctctatgtaacaagctctccacggtctca	3216
O	y	3315	AAACAGGCAGCCATTGTGGCAGTTCTCCAGATGACTTGTCTATGTCCACATCTCGACGTGA	3374
D	b	3217	aaataagcctctctctcttcttgaacactctctgatatctgctctggccacactcttaagcaga	3270
O	y	3375	GTCTCCAGATGATGAGGAGGACTCTCCGACCGCGTGTCTCTGGCCTTGGCGGAATTTTGA	3434
D	b	3277	gggccagagtgagacgaaggagcctctgctcggaaccgagccgacacatccctgaaacaacaga	3336
O	y	3435	T-----GGCCTGGAGGAGATTTGATAGAGCCTTAGGAATACCAGAACTGTGTACGCCAG	3488
D	b	3337	tgcacacgagctctggaaggagatcgacagcgctcttggaatctcttgagctcgatgaatcagg	3396
O	y	3489	CCAAGCGATGTGATTCACAGACAG-----TTCTCACTCAGGATTTCCAAATCATCTGTGGA	3542
D	b	3397	acaagagctctgagatcgcaaaacagatgtcttccaaagcccaagaagcagatgatatgta	3456
O	y	3543	GCACACAGGCGCCGGTTTTTCCACAGACAGATATGATCTCAGGCGCAAAATGGCCACAGGTAG	3602
D	b	3457	tcgaagaagctcgacataatgagacagacataccagatcgaagctctccctcaagaagag	3516
O	y	3603	CTATTCTCCCATCCAGATATCCAACTTTCACACC-----TAGGACACAGCGGCTCAGTTA	3656
D	b	3517	ctttaaactctcaggaagcagccacacatcgtttaaacttaataatgagatgagatgaacga	3576
O	y	3657	TGCCACACTCCGATGCGAGCCGAG---ACCGGGCTCAGGGCCACAGGCGCTTAGTGACAGA	3713
D	b	3577	agcgagcttccctctcgaaagcagcatgcatcctagagcgccgtcgtgagaccaaggaacca	3636
O	y	3714	CCACCCAAATCACAATTAAGACTTAACTTTCAGATGCGCTCCACAGCACAGCAGAAATCGCCA	3773
D	b	3637	caccccgaaagcagctgagaaatgagatcttaagaacgaaggtctacag-----ggca	3684
O	y	3774	GCCACTTATCAATCAATCAATCAATGCAATTTTCCAAATGGAATGGAATCTGACCTGAGGCGCGAGT	3833
D	b	3685	gcagattttaaactcagagacggcgagcgacttgaaatgaaataatgysgaacccctgctgagac	3744
O	y	3834	ACCAACACAGGACCTATTAAATGCACAGATGCTGGGCCACAGACAGAGGAAATCTTGAA	3893
D	b	3745	tgcgtgtagaagcccatgctgcgcagagcttctttaaagcccaatgctgctgcccagca	3804
O	y	3894	CCAGCATCTTCGACACAGACAAATGCATGACACACAGCAAGTTCACAGCAACATCTTGAT	3953
D	b	3805	gaaacggaagctgatatgcatctaacctctgacagacgaagatctgcatgatatgtcac	3864
O	y	3954	GATGAGAGGACAGAGGGTTGAAATATGACACCAAGCATATGGGCTCTAGTGGATATGCCAGC	4013
D	b	3865	accacagccctcagagcctctcagcccacactcccaacgtcaecgcctctcccagaatgagcgg	3924
O	y	4014	AACATGAGCAACCCCTCGGATTTCCCAAGGCAAAATGCACAGCAGTTTCCATTTCTCCAAA	4073
D	b	3925	ggttttgacggtctcagcaatgctgcaagccctccacaacagtttccatataccagaa	3984
O	y	4074	CTACGGAAATAAGTCACGAACATCTGATCCAGGCTTTACTGGGGCTACGACTCCCCAGAGCC	4133
D	b	3985	ttacagaaatgggacaacacacagacgacgacttttgtctgaagctctgaaatctccacgtgc	4044
O	y	4134	ACTTATGTCACCCCGAATGGACATATCACAGATGCTCCATGATGACAACAGTCTCAGGCCAA	4193
D	b	4045	aatgtagtctcaagaatggggtcttccagaatgcatatgctgacgaatctctcagccac	4104
O	y	4194	CCGAGCCTATCAGGCCCCCTCGACATAAATGATGGGCGCGAGGGGAAACATGGCGGAAA	4253
D	b	4105	accacatgatcag-----cttcagataatgaagggtggtccgtcagggaacctgycagga	4161
O	y	4254	CAGATATTTTCCACAGCAGTCCCCACACACATTTGGGCGACAGCAACACACAGATGTA	4313
D	b	4162	tgggtctctcccccaggaagatttgctctcccaagggaaacccctgtagctctcaaatatgt	4221

[illegible]





QY	2543	AGCCTGTTGACCAACCTCTGGCAAGTGAAGTGCACAACTTTGGAGAGATTTTGGATGATTTTGC	2602
Db	2448	agaacgaacgaagagatctcgagagacctgagataactagatgcactcttggaaacttcttga	2507
QY	2603	AGAAATAGCAATTTTCCACACTTTTCCACACACGAGGGCCAGGGCCCTCGTGGATACAG	2662
Db	2508	ccagttctcgacttctacaacaactccacaalctgacggttaccaccagggggccaaacagag	2566
QY	2663	TTTGACAAAGCAAGCCATCATCAATGACCTCTGCAACTCTGCACACTGTGAAAACAGCCCTGTCA	2722
Db	2567	atgttttcagagaccgagttctctcgggttttcggaagtcacacag-----ctgtgc	2615
QY	2723	CACCTGTTGGAGCCCAAGAAAACAGCACTGCGAAATTTCCAGAGACACTTTTAATTAACCCAC	2782
Db	2616	agttgttctgcctccataataaccgagcggtgtctctgtagtaaccctgtgtctgtgc	2675
QY	2783	GACCAAGGCAACCTGGCGAGGTTATTTGGCAAAACAGAAATTTACCCTCTTGACATCACTATTC	2842
Db	2676	caagttccgcgcagtgaaagatgtcagttgtcttccctgggttaccaaaacagcccatctcgg	2735
QY	2843	AAACCCCAACTGGTGCTGGAGACTTTTCCACAACTCAAGAAACAGTATGTCCTCTCTAGGA	2902
Db	2736	cttgggaatccaagaatgtatgtatagtcagagaaatlaagtcgtccaaatctgggcccaca	2755
QY	2903	TACCTACGACCAAGATGATGGGTATTCAGAGGATGTAGAAACCAAGAAATTTTAGGA	2962
Db	2796	gaatgtcttcgttgbaatccgaactctcccccggagactctgggtcttagtctaactaagg	2855
QY	2963	ACACTTACACAGGATATGTTGGTAAACAGTGTCTTCGGCTTACTATGCCATCTGGAGAT	3022
Db	2856	ccacgaagaatctggagcctctcgtgcatacaagttccctctggaaagactctggagcgaatcagtg	2915
QY	3023	GGGACACGCGACAGTTCGGCTGTAGAGATCACTGTCTCTCAACCAAGTGCATGAAC	3082
Db	2916	ccacttaccagacagctcgtccatctgggggc-ctgtgtccactcttgccacttcgttctat	2974
QY	3083	GGCCAGTCCAGGAGGTATGATTTCCGAAACCCAGACCCAGCATCCCATGAGGCCACCA	3142
Db	2975	cgactgtccagtgatgaagaccactgtgtcagcaacag-cagcagcaacagcagcaacag	3033
QY	3143	GGCAGCCTGGC--CAAAAGACAGACGCTTCAAGTCTCAAGTATATAGGGCCATCTCGA	3200
Db	3034	acaacaaacagcaagaagaagaagaagaagaagaagaagactgtcctaalgaaagaaactcgtga	3093
QY	3201	ATTAGAAATCAACATGGGGGAGCCTCATGATATGCCAACAACAAGCTCTCCAAATCAAC	3260
Db	3094	gattcccatctgggaatctgggaatcactatcactatgcca---gcagtgcagttactaaccaacc	3150
QY	3261	TGCCCCATGSCCTGAAAGACATCTCGCTATATAGCAAGCGGCTTTTGGCAGCAAAACAG	3320
Db	3151	agttctctcgcagagggcatalgtctctctatgtgaacaagttctctcaggtctccaataatg	3210
QY	3321	GCACACCTTTGGCAATTTCTCAAGATATACATTTGCTATGTCCACATCTCTGCAGCTGAGTCTCC	3380
Db	3211	gcccctctctagaaacctctctgtgatgtatcgtcttggcgcaacctttaaagcagagggcca	3270
QY	3381	GAGTATGAGAGGAGCTCTCTGTGAGACAGCTGATATCGGCTTTGGCGAATTTTGTAT----	3435
Db	3271	gaagtgaacgaagagggcctctcgtctgtgaacaagcttgcacaactctctgtgaacaacaagatggcac	3330
QY	3436	-GGCCTTGAGAGATTTGATAGAGCCCTTAGGAATATCCCAACTGCTCAACGACAGCCAGC	3494
Db	3331	aggtcctcgtgagagatcgcagcagggccttgggaattctcgtgcctcgtgaatcaagggaacaagc	3390
QY	3495	AGTATGATCCGAACAG-----TTCTCAAGTACAGATTTCAACATCATGCTGGACGAA	3548
Db	3391	tttggatgtccaaacagatgttttccaaagggccaagaagcagatgaatgatgtatgcatcagaa	3450
QY	3549	GGGCGCCGTTTCCCAACAGCATATCATCTCAAGGACAAATAGGCCCAAGGGTACGTATATTC	3608
Db	3451	ggctgcactatatagtgacagacatacccaagctcaagggtctcccttcaaggaggtcttaa	3510
QY	3609	TCCCATCAAGATCCAAACTTTTCACAC-----ATGGACACGCGCGCTACTGTATGTCAC	3662

[illegible]

XX WO9857982-A2.  
PN  
XX  
PD 23-DEC-1998.  
XX  
PF 17-JUN-1998; 98WO-US12689.  
XX  
PR 17-JUN-1997; 97US-0049728.  
XX  
PA (USGO ) US GOVERNMENT.  
XX  
PI Meltzer P, Trent JM;  
XX  
DR WPI; 1999-080946/07.  
XX P-PSDB; AAW81025.  
DR  
XX  
PT New isolated steroid receptor co-activator, ARN1 - used to develop  
PI products for the diagnosis and treatment of steroid-responsive  
PT tumours, e.g. breast, lung, prostate or colon cancers or melanomas  
XX  
PS Claim 7; Page 25-30; 57pp; English.

The AIB1 protein is a member of the steroid receptor coactivator-1 (SRC-1) family of nuclear receptor co-activators that interact with estrogen receptors (ER) to enhance ER-dependent transcription. The AIB1 gene is amplified and over-expressed in certain cancers in particular breast cancer and steroid hormone responsive cancers. The AIB1 polypeptide can be used to identify compounds which inhibit ER-dependent transcription. Increased expression of the AIB1 gene indicates aberrantly proliferating cells, thus detection of increased expression of the AIB1 gene or an increase in the number of copies of the AIB1 gene can be used to diagnose cancer or a predisposition towards developing cancer. Compounds which inhibit expression of AIB1 or compounds which inhibit interaction of AIB1 with steroid receptors or nuclear co-factors can be used for reducing the proliferation of cancer cells.

Sequence 6035 BP; 1988 A; 1479 C; 1528 G; 1840 T; 0 other;

Query Match	8.88;	Score 542.8;	DB 20;	Length 6835;
Best Local Similarity	51.58;	Pred. No. 6e-140;		
Matches 2241; Conservative	0;	Mismatches 1762;	Indels 351;	Gaps 29;

0Y 141 CAGTGGCTGATATGTGTTCAAGATGAGTGGGATGGGAAAAAATACTTCGACCCCTCCAG 200  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 179 caagtcgcgatgatatattccaaagtatgtagtttagggaga---cttgatacctaacyc 235  
  
0Y 201 GGCAGACACAAGAAAGCGCAAAGAATGTCTTGACCAACTTGGACCACGCCAAAAGGA 260  
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 236 cagtgattcacgaagaacgcaatttgcactgtaaccacgagaacgaccttaacttcag 295  
  
0Y 261 CACCTAAAAACGTAATCGTGAACAGAAAAATTAATATATAGAAACTTCGACAGTTGAT 320  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 296 tggtagaaaaacgagacgcggagacgagaaaagtaatacatattgaagaaatttgcctgagctgat 355  
  
0Y 321 TTTTGCAAATTTTAATGATATAGAACACTTTAACTTCAACCTGACAAATGTGCATCTT 380  
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 356 atctgcacaatcttagtgatatttgacaattlccaatgtccaacacgataaatctgycgatlft 415  
  
0Y 381 AAAAAGAACTGTGAAGCAAAATTCGTCAGATCAAAAGAACAGAGAAACACCACTGGCAG 440  
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 416 aaaggaaacagaaagacagatacgcaccaataaagaagacagaa---aacattlccaa 472  
  
0Y 441 CATGATATGAAGTGCAGAGAGTCAGATGTATCTCTACAGAGGCGAGGTGTCATCACAGA 500  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 473 tgatgatgatgttccaaaaaaccccgatgtatctcttccacagggcaggyagtatattataaga 532  
  
0Y 501 TGCGCTGGGGCTTATGAGACTTGAAGCCCTTGATGGGCTTCTCTTTTAGTGAACCTGGA 560  
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 533 ctcccttagaacgccttctaacttcagcgaatttgat-gtgttctctatttbglttgatctogaga 592  
  
0Y 561 AGGCAACCTGTGTTGTTGTGCAGAGAAATGTACACAGTATCTTAAGTATTAACCAAGAGA 620

Db	533	cggaaacatttgatatttgatcagaataatgctcacacatacctcgcaataataagcagaagga	652
Qy	621	gctgataacaaaagtgtatatatagcatcttgcatgtgtggggacacacacgaattttgtcaa	680
Db	653	cctggttaacacaagtgcttaccatattcttaccatagagaagacagaagaatttcttaa	712
Qy	661	aaacctgctgcgaagctctatagtaaatgggggactgttgctgcgcgaacctccagcgcg	740
Db	713	gaa---tttaccaaaacttaccagttatgagtttccctggaanaatlgagacccaagaca	769
Qy	741	gaacagccatctccttcaattgtgcgaatgcgtggaataaccttacctgatttcagagaga	800
Db	770	aaaaagccatcatattatttgcgttgatgtgtaaaaacaccacatgatattcttgyaaga	829
Qy	801	gggttcattgataccaggaagctcatgaaatvtgaaatcatgcagtgcttgcgtgc	860
Db	830	cataacgcgcagctctgtaaatgctgcagagatatagaaacatgcaagtgtcttgccttc	889
Qy	861	tcacaccaaagctccattcAAAAGAGAGAGAGATTTCGAGTCTCGTTGATTTGGGTGGC	920
Db	890	tcagcccaagagctctatgtaggaagaagggaaatttgcacattgtaatctgtaatctgtg	949
Qy	921	aagaagacttccatctgaagaaaagacacagctcttccctcatgagaaaagtttactacgc	980
Db	950	acgcgcattactacagagagaagaacatttccataaacacctgtagagattattaccag	1009
Qy	961	ccagatctccaaaggcaagatcACAGTCTGTGATACcAGcAcCATTGAGcAGccATGAA	1040
Db	1010	acatgatcttcaggaagaagtgctcaatatagatacaaatlcaatcagtagtccctcatag	1069
Qy	1041	acCAGcGTGGGAGAGAcCTGGTAAAGAGcGTATTTCAGAGTTCATGcCAGAcATGAGG	1100
Db	1070	gcctgagcttgtaagataataatccgaagtgtagcttcagaaattttagcttaaagtatg	1129
Qy	1101	AGAACTCTGTGCTCCTATGCTAAAGAGcCTCATCTGAGAGTACAGAcAGATTTCGCA	1160
Db	1130	gcagctatggtccca---ggaagctcactatacgaagattatcttaatgscatagaca	1166
Qy	1161	CAGTCAATCTATCTGTTTTTCCTTGTCTGATGGCAGCTCTTGTGCTGCACAAAGCAGAG	1220
Db	1187	aaacctatatactgattctcgttgtagtgaactatagtagactgacagacagaacaaaag	1246
Qy	1221	CAAACTATCCGCTTCAGAGCTACTATAGAAcCTCAACTGTAATTAATCTTTACATAGCT	1280
Db	1247	caaacctctccgaaatccgttaaaaaatgtagcagtgttcttccaacccacttct	1306
Qy	1281	TCACAGAGcAGAGATGTGTGTGATGAATCCGAGTCTGAGTGCAGCAACAGATGGGAA	1340
Db	1307	tcagagagaaacagatagatataagacaaacccaatctcgttgscagaagatattagcc	1366
Qy	1341	gccactgattccaatttagcttTAACAGcCTGCCCATCAGAGcCCTGTGcATGtGGGAACC	1400
Db	1367	acctatggc-----tgaatgcaacagctcgtagagcgcatagatgltgcacaaaca	1420
Qy	1401	AGGTcAGAGcATGAGcCCTCAGTGAACATATAAATTTTCCCATAAATGGcCAAAAGACAA	1460
Db	1421	aggcttacaagatgcgcagacacagggcctatggtcgtgcagacacctagacacaagggca	1488
Qy	1461	AATGGcATGcCCATG---GGCAGGTTTGcGTTCGGGGcAGATGACCATGTCTCAGG	1517
Db	1481	gatggtgagctagatgtagtggggttccagtaacatagcttcatbgaaccttggcagg	1540
Qy	1518	CATGCAACACACACTGC---TCAGGcTAACTATCTGCAcTCAATTAATGAACAGcCCTC	1574
Db	1541	catgcaatcacatcttcttaccagaacaaacactatgggtctcaacaatagaaagaccccc	1600
Qy	1575	ACAAAGcAGcCCTGGcATGATCCAGAGcAGcCCACCTCAGTCTTTACCAAGAcATCG	1634
Db	1601	acatbgyagatccttggttctgcacaaacacgcagcaatatacatttcttccctgtaatcg	1666
Qy	1635	CATGAGcCCTGGAGTGCCTGGCAcCCTCGAATCCACCAcAGTCAcGTTTTCCTTCAGG	1694



Db 1661 -----tggagtlccaagaatagatccatcatgttctctgttc 1702  
QY 1695 AAGCTGCAATTCCTCGTGGAGTTTGACAGACAGAAATGCAATGTTATTACCA 1754  
Db 1703 aggtgtgacatctccatcgcatctcttgcgaatactgga-----ccacgcttccag 1759  
QY 1755 CAGCTCCCTCAATGCACTTGAAGCCCTCAGAGAGGGGACAGGGGTTCATTAAGGTCATC 1814  
Db 1760 cagctctcagtcagtcctgcagcatcagtggaagtggtggagcttcccttatctac 1819  
QY 1815 GTTGGCTTCAACAGACTTAAATATGGCAATTTGCCAAATCTCCAGTTAATGAATCC 1874  
Db 1820 tctgtctacccagccccaattggaatactctccc-----aatalgaatat 1867  
QY 1875 TCCCCCACTGACGAAGTGGGAAGCTTGACATCAAAAGACTGTTTGACATATGGGGA 1934  
Db 1868 taccacaacgaataagtaagcaatcagatccaaagatctctgtgcttctatctgcga 1927  
QY 1935 GCCCTTGAAGGTACAACTGGACACAGAGAGCAGCTGCCATCCTGAGACAAAAGA 1994  
Db 1928 ccaaatccagtgagagtgatcaatgtgtcagtaataagcaagatcaacctcagtgaca 1987  
QY 1995 AACAAATGACCCCACTGCCCCCGCGGTGAGCATGAGAGCTGACGGGACAGCAG 2054  
Db 1988 agaaagtaa-----ggagagcagtggttgaaggagaggaatacaagg 2032  
QY 2055 ACTGCATGACGACAAAGGAGACCAAACTCTGACACTGCTGACCAACCAATGTCATCA 2114  
Db 2033 tctcttggaaagcaaaagtcataaataactcagtaacttacttcttcttctgtga 2092  
QY 2115 GATGAGACCCCTCGCCCTTACCAAGCTTTGTGCGATACAAAC---AAAGACTCCACAG 2171  
Db 2093 ccggggatctctctctgtgcacaactcccccagatgataagtgtaagaatcttctgt 2152  
QY 2172 TACCTTGCCCTGG-----TTCGTG 2189  
Db 2153 tagtgtacacagccctctgagatctctctctacatctggaaggtatctctctacatc 2212  
QY 2190 GTCTACACATGGAACCTCGTCAAGGAGAGCAATAAATTTGACAGACTCTTGACGA 2249  
Db 2213 caataatgcatggtctactgttacaagaagcaccggaatttgcacaagtgtgcgcgaa 2272  
QY 2250 CACAGACTCCCTGTGAGCTTGGCAATTAAACAGCAGAACCCACAGCAAGACCTGAG 2309  
Db 2273 tgggaattccacagctgagtgagtaagatactcagagaagccactggtgaagacacag 2332  
QY 2310 CAGAGAGTCCAGACGACACTCTGTGATGAAAGTACTTTTAAACAGAGCCGCTGAG 2369  
Db 2333 -----cagataactcttctgtgggaaggaatgtgtcacaagcagagcagctaa 2383  
QY 2370 CCCCAAGAGAAGAAG---AATGCACTACTGCTGATTTGCTGATTAAGATGATACTAA 2426  
Db 2384 tcttaagaagaagaataatgcatctctagataccctgtgcagcaggtatctctcag 2443  
QY 2427 AGATATTGTTTACAGAAATAACCCCAAACTTGAGAGACTGACAGTAA---GACAGA 2483  
Db 2444 tgtgtcactctctaagaactacagcccgaagtgggaagtggtgaataataatgagctca 2503  
QY 2484 TCTGCGCAGTAACACAATAATTAATAGCAATGAAAACCTGACAGAGAGATGACCTTTGA 2533  
Db 2504 gtgcacacagctccacacttctcagtaagtaagaagaacccctlaaataagacaga 2553  
QY 2544 GCTGTGAGCAGCTGTCAGTGTGAGTGGACAACTTGAGAGAGATTTGATGATTGCA 2603  
Db 2564 gacaagtgaagaagagatctgtgagacttgatatactagaigtatctctgtgtgcagac 2633  
QY 2604 GAATAGTCAATTAACACAGCTTTTCCAGACAGAGGCGCCCTGCTGTGATCAGT 2663  
Db 2624 tagtctgacttttaacaataatccata-----tctcaaatgagtagt 2666  
QY 2664 TGACAGACCAAGCCATTCATGACTCATGACACTACAGCTGAAAAACAGCCCTGTAC 2723  
Db 2667 catctgggagactaagcacagtggtttcaaggaactaa----- 2704

QY 2724 ACCGTTGGAGCCCAAGAAACAGCACTGCGAATTTCAACAGAGCACTTTTATAACCCAG 2783  
Db 2705 ---tctctgtgttgaagaatctacagctctgtgcagctctatctgtccctcataaacg 2761  
QY 2784 ACCAGGCAACTGGCGAGTTATTGCCAAACCAAAATTTACCACTTGACATCATGTGCA 2843  
Db 2762 agcagtgctctgtg-----a 2776  
QY 2844 AAGCCCAACTGCTGTGACCTTTCCACCAATGAGAAACAGTGTCCCTACTGATGAT 2903  
Db 2777 tagccctgttctgtgtgtcgaagctcccgtaaaaaatactcagtgcttctcccaagctt 2836  
QY 2904 ACCTCAGCCAGGAATGATGGGT---AATPCAAAGGATGATGGAAGAACAGGAATTTAGG 2960  
Db 2837 accaaagcaaccatgltgtgggtggaatccaaagatagtatgatacaggaataatcag 2896  
QY 2961 GAAACAGTAGCACGAATGATGGTAAACAGTCTTCGCGCTACTATGCAATGAGGA 3020  
Db 2897 ctcaagtaatggtgtgtgcacaacccgaatgtgactgtactcagactctctctcagga 2956  
QY 3021 ATGGGCAACCGAGATTGCGCTGTGAGATCACCTGTGCTGTACACAGTGCATGAA 3080  
Db 2957 ctgggtcttaccacaactcaaaagcgcgcaagatgtgaactatgaaatcactcagtg 3016  
QY 3081 CCGGCCAG-----TCCAAAGAGGTATGATTTGGGAACCCAGCAGC 3119  
Db 3017 aagaacgagaggaatataataacttcttaccagaccctgactgtgtgtatctcc 3076  
QY 3120 CAGCATCCCATGAGGCCACAGCCACGACCTGCGCAAGAC-----AGACGCTTCAGTC 3173  
Db 3077 caactgtccctctgtgtcttaataagcatataccaggtgtgagaccagatgttgcacgcagca 3136  
QY 3174 TCAGGTATGAAATPAGGCCATCTGAATTAGAGATAAATGGGGGACCTCATPAG 3233  
Db 3137 gcaagatgtcttaaaatgagcctgtgtgaatcccatgtgaatgtgggtctaaatccctatg 3196  
QY 3234 CCACCAACAGCTCCTCCAAATTCAGACTGGCCCTGGAAGCAATCCGCTTATAGA 3293  
Db 3197 ccaagcagcagcatct---aaccaatgtgtctctgtgcgaatgtgtgtccatgga 3253  
QY 3294 CAGGCGCTTTTGGCCAGCCCAAAACAGGACGCACTTTGGCAGTTCTCAGATGACTTGT 3353  
Db 3254 acaagttctcattgtgcactcaaaatagccctctctttagaattccctgtatgtctgt 3313  
QY 3354 ATGTCCATCTCTGACACTGACTGCTCCGAGTGTATGAGGAGACTTCCGTGACACGCTGTA 3413  
Db 3314 tgggcccaccttccaaacctgtgaagccagagtgacgaagaagcattatgtgacacagctgca 3373  
QY 3414 TCTGCGCTTGGGGAATTTTGAT-----GGCTGAGAGAGATTGATGAGGCTTATGGAAT 3467  
Db 3374 cactctctcagcaacaacagatgtccacagacctgtgaagaatgtacagagcttgtggcat 3433  
QY 3468 ACCCGAAGCTGTGACGACAGGACCAAGCATAGATCCAGAACAG-----TTCCTCAAGTCA 3521  
Db 3434 tctgtgaactgttcaatcaggaagcagcatlaggcccacaacaggaatgtcttccaagcca 3493  
QY 3522 GGATTTCCAAATCATCTGCTGAGCAGAAAGCGCCGTTTCCACACAGCATGTGATGCA 3581  
Db 3494 aagaagcagcagtaatgtatcagaagcaggaattatgtgacaagacataccacagaca 3553  
QY 3582 GGCACAAATGGCCAGGCTGATTTCTCCATGCAAGATCCAAATTTTCAACACATGGG 3641  
Db 3554 ggggcccacaaatgcagaagagcttctatcttcaaggaacaatcacatctttaaactat 3613  
QY 3642 ACAGCGGCTAGTTATGCCACACTCCGTATGCAAGCCAG-----CCGGGCTCAG 3692  
Db 3614 gatgaatcagatgaacacaggaagcttctctctcacaaggaatgtacccaagagcaaa 3673  
QY 3693 GCCACGCGGCTACTGACAGAACCAAGCCAAATCAACTTAAGACTTACGATGCGCT 3752  
Db 3674 catcatgagaccccggaacaacacccccaagcaactttagatgtcaggtcagagaggt 3733





```
OY 2844 AAGCCAACTGGTGGAGCTTTCCCAATCAGAAACAGTATCCTTACTCAGTGTAT 2903
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2761 tagccctgttctgtgtgtcgaagtcctccagtaaaatcaccgtgtcttcccatgt 2820
OY 2904 ACCTCAGCCAGGAATGATGGGT---AATCAGGATGATAGAAACCAAGAAATTTAG 2960
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2821 accaaagcaaccatgtgtgtgtggaatccagaatgtgatactgagaaataltatg 2880
OY 2961 GAACGATGACAGAAATATGTAACGTCTTCTGGCCCTACTATGCCATCTGGAGA 3020
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2881 ctaaaagtatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2940
OY 3021 ATGGCACCCGACAGATGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3080
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2941 cgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 3000
OY 3081 CCGGCCAG-----TCCAGAGGATGATGATTCGAGACCCAGCAGC 3119
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 3001 aagaccagagagatataactcttctaccagacctgtgactgtgtgtgtgtgtgt 3060
OY 3120 CAGCATCCCATGAGGCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 3173
      || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 3061 caagtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 3120
OY 3174 TCAGGTCAATGAAATATAGGCGCATCTGATTTAGATGAACATGGGGGAGCTCAGTATG 3233
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 3121 gcaagtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 3180
OY 3234 CCACACACAGCTCTCTCCAAATCAGACTGCCATGGCTGAAAGCATCTGCTATTA 3293
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 3181 ccaagcagagatct---accacatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 3237
OY 3294 CCAGGCGTCTTTGGCAGGCCAAACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 3353
      || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 3238 acaagtcttctacgtacccaataatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 3297
OY 3354 ATGTCCACATCTGCGAGTCTCCAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3413
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 3298 tgggcacacttccaaactcgtgaaggtcagagtgacgaagacatattgtgacagctgtca 3357
OY 3414 TCTGGCCTTGGCGAATTTTGAT-----GCCCTGAGAGATTTGATAGACCTTAGAAT 3467
      || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 3358 cactcttctccacacacagatgtccacaggtcctggaagatgtacagacttgtgtgt 3417
OY 3468 ACCCGAAGTGTGACGCGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 3521
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 3418 tctgtgaactgttcaatcagagacagatagagccccaaacagatgtgttccaaggtcca 3477
OY 3522 GGAATTCGAATCATCATCTGTGAGAGAGAGGCGCGCTTTTCCACAGCAGTATGATCTCA 3581
      || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 3478 agaagacagcatatgtatgtatcagaagagatattatgtacagacataccagacaca 3537
OY 3582 GGCACAAATGGCCCGAGGTATGATCTTCCCATGCAAGATCCAAATTTTCCACGCATGGG 3641
      || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 3538 gggggtcctcaatgtcaaggggtgttcatctcagggacatccatctttaaactctat 3597
OY 3642 ACAGCGGCTAGTTATGCCACACTCCGTATGACGCCAGA-----CCGGGCTTCAG 3692
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 3598 gatgaatcgaatgaacacagagcaatttctctccaaggaatgtcacaccagaagaccac 3657
OY 3693 GCCACGGGCTTATGTCAGAACCCAAATCAACTTAAGACTTCAACTTTCAGAGCTGCGCT 3752
      || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 3658 catcatagagaccccgagacaaaccccccaagcaacttagatgtcagcttcaagagagct 3717
OY 3753 CCAAGCAGCAGCAGAAAT-----CGCCGACCACTTATTAATCAATCAGCATGT 3800
      || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 3718 gaaagggccagcaglttltgaatcagagccagagcactgtgaatgtgaataatgtgaaccc 3777
OY 3801 TTTCGAATG-----TGAACCTTGACTGTGAGGCTGAGTACCAACACAGCGC 3845
      || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 3778 taactgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 3837
```

```
OY 3846 ACCATTAATGCACAGATGCTGGCCCA----- 3872
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 3838 tttcttaatgtcctaattgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 3897
OY 3873 -----GAGACGAGGGGAAATCCGTGAACCCAGCATCTTCGACA 3908
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 3898 acaacagaggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 3957
OY 3909 GAGCAAAATGCACTACAGCAACAGCAAGTTCCAGCAAGCAACTTTGATGATGAGAGCAAG 3968
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 3958 gcaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 4017
OY 3969 GTTGAATATGACACCAAGCATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4028
      || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 4018 cttagccaccctcttaatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 4077
OY 4029 TCGGATTCGCCAGGCAAAATGACAGAGTTCATTTCTCCAACTAGCAATAATGTC 4088
      || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 4078 cacaatgtccaaagctcctccgcgaagatltccataccaacaaatltatgtgtgtgtgt 4137
OY 4089 GCAACCTGATCCAGAGCTTTACTGTGGGCTACGACTCCCGAGAGCCCACTTATGTACCCG 4148
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 4138 acaaccagatccaagcttgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 4197
OY 4149 AATGCGCATPACACAGATGCCATGTATGCAACAGTCTCAGGCCAACCCGCTATCAGCC 4208
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 4198 aatgtgtcctccccaatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 4255
OY 4209 CCCCTCCGACATAAATGATGGGCGCAGGGAACATGGCGGAACAGCATGTTTCCCA 4268
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 4256 -tccatagaagaatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 4314
OY 4269 GAGTCCGCCACCACTTTGGGCGAGCAAGCAACACACAGCAGT 4311
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 4315 gcaagcagttgtccaccaggggaatcctgtcagtgatagtag 4357

RESULT 6
AAF26490
ID AAF26490 standard; DNA; 6754 BP.
XX
AC AAF26490;
XX
DT 27-MAR-2001 (first entry)
XX
KW Steroid receptor coactivator-3; SRC-3; antisense; infection;
XX inflammation; tumour; cancer; ds.
XX
OS Homo sapiens.
XX
PN US6156571-A.
XX
PD 05-DEC-2000.
XX
PF 15-NOV-1999; 9905-0440612.
XX
PR 15-NOV-1999; 9905-0440612.
XX
PA (ISIS-) ISIS PHARM INC.
XX
PI Bennett CF, Cowser LM;
XX
DR WPI; 2001-079549/09.
XX
PT Novel antisense compound useful to prevent or delay infection,
XX inflammation or tumor formation, specifically hybridizes with and
XX inhibits the expression of human steroid receptor coactivator-3
PS Example 13; Column 43-54; 36pp; English.
XX
CC The present invention relates to an antisense oligonucleotide.
```

CC targeted to a nucleic acid molecule encoding human steroid receptor coactivator-3 (SRC-3). The invention is useful for inhibiting the expression of SRC-3 in human cells or tissues *in vitro*. It is useful for diagnostics, therapeutics, prophylaxis and as research reagents and kits. It is useful prophylactically, to prevent or delay infection, inflammation or tumor formation.

XX Sequence 6754 BP: 1955 A; 1463 C; 1511 G; 1825 T; 0 other;

Query Match 8.3%; Score 510.6; DB 22; Length 6754;  
Best Local Similarity 51.1%; Pred. No. 5.8e-131;  
Matches 2219; Conservative 0; Mismatches 1769; Indels 357; Gaps 29;

```

QY 141 CAGTGTGATATGTTGATGATGATGGGATGGAGAAATACCTGACCCCTCCAG 200
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 162 cagttgcgagatgatalatcaagaatgagatgaagaaa--cttgcgaccactgac 218
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 201 GGCAGAGACAAAGAAAGCCCAAGGAATGTCCTGACCACTTGGACCCAGCCCAAGNA 260
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 219 cagtgattcaagaagaacgaatgcatgatactccagacagaatgcttactctgacg 278
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 261 CACTGAAAACGTAATGCTGAACAGAAATATATATATAGAGAATCTGCAGACTGAT 320
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 279 tggtagaaaacggagacggagcaggaagaatataatattggaagattgctgagctgat 338
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 321 TTTTGCAATTTTATATATATAGACACTTTACTTCAACCTGACCAATGTCGCAATCT 380
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 339 atcgcgaactctagatgataatgacaattcaatgcaacagaataatgycgactt 398
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 381 AAAAGAACTGTGAAAGCAATTCGTGACATCAAAAGAACAGAGAGCAGCTGCCAA 440
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 399 aaagaaacagtaagacagatacgcacaataaagaagaagaa--aacatttccaa 455
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 441 CATGATGATGTCAGAAATGATGATGATCTCTACAGGGCAGGGTGTCATGACAGGA 500
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 456 tgaatgatgtctcaaaaagccgatatcttctacagggcagggagattatgataaaga 515
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 501 TGGCGTGGGGCTATGATGCTTGAGGCCCTTGATGGGCTTCTTCTTGTATGATGACCTGA 560
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 516 ctcttaagaccgcttacttactcaaggcatggaatggttccctattggtgaaatcgaga 575
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 561 AGGCACCTGTTGTTTGTGTGAGAGAAATGTGACACAGTATCTTAAGGTATTAACCAAGA 620
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 576 cggaaacatgattatgataagaataatgtaacacaataccgcaataataagaagaaga 635
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 621 GCTGATGACAAAGCTGATATATGATCTTGATGATGTTGGGACACACGGAATTTGTCAA 680
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 636 cctggtatacaagaatglttacaatatcttacaatgaagaagaaagatttcttcaa 695
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 681 AAACCTGTGCAAAAGTCTATATGTAATGAGGGGATCTTGTCGTGGCAACCTCCAGGCG 740
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 696 gaatttac--caaaatctacagttcaatgaggttccctggaacaaatggaacccaagaaca 752
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 741 GAACAGCATATCTTCAATGTCGGATGCTGGTAAACCTTACCTGATTCAGAGAGA 800
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 753 aaaaagcacaataattatgctgcaatgtagtaagaacacacacatgataatcttgaaga 812
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 801 GGGTCATGATTAACAGGAAGCTCATCAAGAAATGAAATATGCAAGCTTCGCTGCTC 860
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 813 cataaacgccagtcctgacatgctgcagagatatgaacaatgacagttctgcccctgc 872
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 861 TCAACCAAAATGTCATCAAAAGAAAGAGAGATTTGACGTCCTGCTGATTTGGTGGC 920
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 873 tcaagcagagatgataaggaaggaaggaagatctgcaatctgtaagatctggtgagc 932
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 921 AAGAAGACTTCCATGAAGAAGACAGTCTTCCCTCATCAAGAAAGTTTACTACTCG 980
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 933 acgcgcattactaacaaggaagaacaattccatcaaacctctgagagcttataacag 992
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 961 CCAGAGATCTCAAGGACAGATCACGTCCTGATACGACGACCATGAAGCCATGAA 1040
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 993 acatgatcttcaaggaaggtgtcaatatagatacaaatcactgagatcctcatgag 1052
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

QY 1041 ACCAGCTGAGGACCTGCTGTATAGAGGTATGTATCAGAGTTCCATGGCAGCATGAAG 1100
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1053 gctctgcttgaagataataatccgaagtgatctcaagatttttagtctaaatgag 1112
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1101 AGAATCTGTGCTTA-----TGGTAAAGGACATCATCA 1133
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1134 TGAAGTACTGAGACAAAGATTGGCATTCAGTCAATATCATGTTTCTTCTGATGG 1193
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1173 aacagcttatacttaatggtccatgacgaagaacccagatatacgtatcgttgcgtgag 1232
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1194 CACTCTTGTGTCGACAAAGCAAGCAAACTCATCCGCTTCAGACTACTCAATGAAC 1253
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1233 aactatagactgcacagacacaaaagaacctctccgaacctgttaacaaatgactg 1292
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1254 TCAACTGTATATATCTTACATATATGCTTCACAGAGAGCAAGATATGTTGTGATGAATCC 1313
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1293 aatggtcttgcctcaacccactctctcagagagaagaatgatalagaccaaacc 1352
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1314 GATCTGACTGACAAAGATGGGAGAGCCACTGAATCCATTTAGCTTACAGCCCTGC 1373
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1353 aaatcctgttgacaagagattagaccactatg-----gttgatgcaacagttcg 1406
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1374 CCATCAGGCCCTGTGTCAGTGGGAAACCCAGGTCAGGACATGACCTCAGTACGATATATAA 1433
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1407 aggcggcatgagatgltgcgccaaccaaaggtctacagatgctgagacagaagccatg 1466
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1434 TTTTCCATTAATGSCCCAAAGAGAAATGGGCAATGTCCTATG--GGCAGGTTGGTGG 1490
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1467 ctgtgcaagacctagacacaaagcagatgagtgagctgagtgatggtgtccagtaa 1526
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1491 TTTGGGGAATGAACATATGTGTGACGATGCAACGACCAACTCTCT---CAGGGATGTA 1547
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1527 catagcttcatgacccttgcgagcagcatgcaatccacttctcactcagaacaa 1586
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1548 CTATGCACTCAAAATGAAGCAACGCCCTCACAAAGACGCCCTGGCAATGACAGGACACC 1607
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1587 ctatagctcaacatgagtagccccaatgaggtgctgtgcttgcgccaacacga 1646
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1608 CACTCTCATGCTTTCACCAAGGACATGCGATGAGCCCTGGAGTGCGGACGCCCTCGAAT 1667
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1647 gaataatgatttctctctgtaatcg-----tggaggttccaaagat 1688
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1668 CCCACCCAGTCAATTTTCCCTGACAGAACTTGATTCCTCTGTGGAGTTTGCAGCAG 1727
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1689 agccacacatcagtttctctgtgtagtgatgcatctccatgcatcttctgcaaa 1748
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1728 CACAGGAATAGCCATATGATTAACCAACAGCCTCTCAATGACATTCAGGCGCTCAGGGA 1787
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1749 tactgagaa--ccaagcttctccagcagctctctcagttgcccgtgcaagcatcagta 1805
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1788 GGGGACGGGGTCTATTAGGCTCATCGTTCCTGACACAGCTAAATAATGGGCAATTT 1847
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1806 aggtgtgaggtactcccttatactctctgctatcacacagccccaatttggataact 1865
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1848 GCAAAACTCCCAAGTTAATATGAAATCTTCCCTCACTGACAAAGATGGGAAGCTTGACTC 1907
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1866 tccc-----aatagaatatcccaacaaagaaagaaagaaagaaagaaagaaagaa 1913
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1908 AAAACATCTGTTTGGACATATATGGGAGGCCCTCTGAAGGTCAACTGACAAAGCAAGAG 1967
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1914 caaagatcctctggtcttatttgagcacaacaaatccatgtagagatctaatgtgtcagtc 1973
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1968 CAGCTGCCATCTTGAGAGACAAAGAAACAAATGACCCCAACCTGCCCCGGCGGTGAG 2027
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1974 aaatagcagagataccaccctcagtgacaagaagaagtaagagagc-----ag 2018
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2028 CAGTGAAGAGCTGACGGGCGAGAGCAGACTGCATGACGCAAAAGGCGAGACCAACTCT 2087
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2019 tgttaggggacagagaaatcaaaaggtgtcttggaaagcaaggtcatataaaatctact 2078
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

[illegible]

Db	3051	gcccttcggtc	taataagc	ataccagc	gtgcgagacc	agattgtc	caacg	-----cagca	3104
Qy	3174	TCAGGTCATG	AATATAGG	CCATCTCGA	TTTGAATTA	AGATAGAA	TGGGGGAG	CCCTCAGTATAG	3233
Db	3105	ggaagtgtc	taataatg	agcgccgtg	agaaatcc	ccatbgtga	atgggggtac	ctccatctg	3164
Qy	3234	CCACACACAG	CTCTCCAA	ATCAGACTG	CCCCATG	CCCTTGAA	AGCATCTG	CCCTATGA	3293
Db	3165	ccaagcagc	agcatct	---aacca	acttggt	ctccgtg	cccgatg	catgctgtgtccat	3221
Qy	3294	CCAGCGCTT	TTTGGCAC	CCCAAAAC	AGGCGAGC	ATTTGGC	AGTTCTCC	AGATGACTTGCT	3353
Db	3222	aaaagttcc	atcctagc	actcaaaa	ataggc	ctctcttc	ttaggaat	ctccgtgatgcttgc	3281
Qy	3354	ATGTCCACAT	CTCGCAGT	AGTGTCC	AGATGATG	AGAGGAC	CTCTCTG	GCACACAGCTGA	3413
Db	3282	tgggcac	cttc	caacac	ccctgga	agcgca	agatgagca	agacatattgtgac	3341
Qy	3414	TCGTGCTT	GGCGAA	TTTTGAT	-----GGCCTG	GAGAGAT	TGATAGAC	CTTAGCAT	3467
Db	3342	caactcttc	cagcaaca	acagat	gtccacag	ccctgtgaa	gaatattg	cagagatcttgggcat	3401
Qy	3468	ACCCGACAT	GTGGTAC	CCAGAGC	AGAGTGA	TGATCC	AGACAG	-----TTCTCAAGCA	3521
Db	3402	tcctgaact	gtc	atcagc	agcagc	atlagg	cccaacag	gatgctctccagagcca	3461
Qy	3522	GGATTCCAA	CAATCATG	CTGAGC	AGACAA	AGCGCCG	CTTTTCC	CACAGTATGCA	3581
Db	3462	agaagcagc	agaaatg	atgatac	agaagcag	atataat	atgacag	aaataccagca	3521
Qy	3582	GGCACAAAT	TGGCCCA	GAGGTAG	TCTTTC	CCATGCA	AGATCCAA	CTTTACAC	3641
Db	3522	ggggtcc	caatgca	agaggtg	cttcac	cttc	taagca	atcaatccatctttaa	3581
Qy	3642	ACAGGGCCT	ACTTATG	GCACAC	CTCCGTA	TGAGCCCA	GA	-----CCGGGCTCAG	3692
Db	3582	gatgaatc	agatga	ccagcag	agaaatct	ctctcc	aaagaa	atgcaccca	3641
Qy	3693	GGCCACGG	GCCTAG	TGCAAG	ACACAC	CAAAATCA	ATAAGACT	TTCAACTTCAGCATCGCT	3752
Db	3642	catctatg	agccccg	acaaaac	cccccaag	caactatg	atgacgt	ctccagagagct	3701
Qy	3753	CCAGCAGAC	AGAAATG	CGCCAGC	-----CACTTAT	GAATCAAA	TGAC	3794	
Db	3702	gaagggcc	agcaggt	tttgaat	cagagcc	agcagc	acttgat	tgaaatgga	3761
Qy	3795	CAATGTTTC	CAATGTG	AACTTGAC	TCTGACTG	AGGCTTG	AGTACCA	ACACAGGCAC	3854
Db	3762	tactgtc	gtgtgtg	tgtcgtg	atga	tgatgag	ctatg	atgacgccccaggggttcttaa	3821
Qy	3855	TGCACAGAT	GTGCGCC	CAGACAG	ACAGGGAAT	TCT	-----	3890	
Db	3822	tgtc	caaatgtgtc	gcccaagc	agcagagag	agctgtc	ttaagtc	atcactccgacaacagag	3881
Qy	3891	-----	-----GAAC	CAGCAT	TTTGAC	AGAGACA	AAATGAT	CATACGA	3926
Db	3882	ggtgtctat	gtatgat	gtcagcagc	agcaaca	agagcagc	agcagcagc	agcagcagca	3941
Qy	3927	ACAGCAAGT	TTCAGCA	CGAATCT	TATGATG	AGAGAGCA	AGGGTTGA	TATGACACCAAG	3986
Db	3942	gcaaacag	caacagc	acacagc	agcaacagc	agcaaa	ccccaggtc	cttcagccccac	4001
Qy	3987	CATGTGGCT	CTCTACT	TGTGTAT	GCCCAAG	CAATATG	ACAACCTG	CGATTTC	4046
Db	4002	tgtgtc	gtcttcc	cccaagat	tgatgggtc	tttgc	agagacc	acaatgcaacaggtcc	4061
Qy	4047	TGCACAGAT	TTCATTT	CTCTCCAA	ATAGAGAT	TAAGTAC	AGCAACCTG	ATCCAGCTT	4106
Db	4062	tccgcaac	aggtttc	atataac	caaatat	atgaaatg	gaacaa	acacagatccagctt	4121
Qy	4107	TACTGGGCT	TACGACT	CCCCAG	AGCCCACT	TATGTAC	CCCGAAT	TGACATATACAGAG	4166



Db 4122 tgcgtcagtgctagctccccaatgcatgctgcgtcaagaatgggtccctccagaa 4181  
Qy 4167 TCCCATGATGCAACAGTCTCAGGCCAACCCAGCTATCAGGCCCTCCGACATTAATGG 4226  
Db 4182 tcccatgatgaaacacccgcagctgcaatccatcatag---tcccaagaatgaaagg 4238  
Qy 4227 ATGGGCGCAGGGAACATGGCGGAACGATGTTTTCAGAGCATGCCACCACTT 4286  
Db 4239 ctggccatcaggaatttggccaggaacagctcttccagcagcagtttgcacaca 4298  
Qy 4287 TGGGACGACGAAACACCAACGACATG 4311  
Db 4299 ggaggatccctgcagtgatagtagt 4323

RESULT 7  
AAA09325  
ID AAA09325 standard; DNA: 4660 BP.  
XX  
AC AAA09325:  
XX  
DT 10-AUG-2000 (first entry)  
XX  
DE Human cancer associated antigen precursor DNA, clone NY-REN-52.  
XX  
XX renal cancer: cancer associated antigen precursor; diagnosis;  
KM cytosolic; steroid receptor coactivator; ss.  
XX  
OS Homo sapiens.  
XX  
XX WO200020587-A2.  
XX  
PD 13-Apr-2000.  
XX  
PF 04-OCT-1999; 99WO-US22873.  
XX  
PR 05-OCT-1998; 98US-0166300.  
PR 05-OCT-1998; 98US-0166350.  
XX  
PA (LUDWIG) LUDWIG INST CANCER RES.  
XX  
PI Obata Y, Gout I, Tureci O, Sahin U, Pfreundschuh M, Scanlan MJ;  
PI Stockert E, Chen Y, Old LJ, Jager E, Knuth A;  
XX  
XX WPI: 2000-303774/26.  
XX  
DR Preventing, diagnosing and/or treating disorders associated with  
PT abnormal expression of human cancer associated antigens  
XX  
XX  
PS Claim 57; Page 96-97; 121pp; English.  
XX  
CC AAA09321-45 were isolated by SEREX screening from a renal cancer  
CC cell line 1973/10.4. Homology searching revealed that these clones  
CC correspond to known genes. The present sequence has identity with the  
CC steroid receptor coactivator gene. The genes encode cancer associated  
CC antigen precursors. These products are useful in methods for  
CC preventing, diagnosing and/or treating disorders, especially cancer,  
CC associated with abnormal expression of human cancer associated antigens.  
CC The method comprises contacting a sample from a subject with an agent  
CC that specifically binds to the nucleic acid molecule or expression  
CC product (or fragment) complexed with a human leukocyte antigen (HLA)  
CC molecule and determining the interaction between the agent and the  
CC nucleic acid molecule or the expression product as a determination of the  
CC disorder.  
XX  
SQ Sequence 4660 BP; 1447 A; 1152 C; 995 G; 1066 T; 0 other;

Query Match 5.9%; Score 364.6; DB 21; Length 4660;  
Best Local Similarity 60.2%; Pred. NO. 2.6e-90;  
Matches 689; Conservative 0; Mismatches 429; Indels 27; Gaps 4;  
Qy 155 TGTCAAGATGATGGATGGAGAAATATACCTCTGACCCCTCCAGGACAGACAAGAA 214

Db 187 ttccaacatgagtgagtgctcggggagagttcatccgaccctcccaaccagaccacata 246  
Qy 215 AGCCGAAAGAAATGTCCTGACCACTGGAGCCAGCCCAAAAAGAACATGAAAACGTA 274  
Db 247 agaggaagaagtcgccc-----atgtagcacatcgtgcatacgaacggaagaagggc 297  
Qy 275 ATCGTGAAACAGGAATAATATATATAGAGAACTTGCAAGATTTGATTTTTCAAATTTTA 334  
Db 298 gcaggagagcaagaataataatattagagaacatcagctgagttactcgtcgcacata 357  
Qy 335 ATGATATAGCAACTTTAACTTCAAACTGCAATGTCGCAATCTTAAACAACTGCA 394  
Db 358 gtgacatgacagcttgagtgatgtaaaccaagcaaatcgaagatttgaagaacaagctcg 417  
Qy 395 AGCAAAATTCGTACATCAAAAGAACAGAGAAAGAGCAGCTGCCACATGATGATGTC 454  
Db 418 atcagatcacgtatagagaatgagaacaagaagaatacaacatcgtatgacgtatgac 477  
Qy 455 AGAAGTCAGATGTCCTCTACAGGCGAGGCTGCATCGACAAGATCCGTGGGCGCTA 514  
Db 478 agaatacagacatctcatcaagtagtagaagaagtagaagaagaatccttggaccgcc 537  
Qy 515 TGATGCTTGAGGCGCTTGATGGTCTTCTTCTTGTAGTGAACCTGAAAGGCAACGTTGT 574  
Db 538 ttcttttggagcttggatgagtttcttcttctgttggaaactggaaggagaattglat 597  
Qy 575 TTGTGTCAAGAAATGTGACACAGTATCTAAGGTATTAACCAAGAAAGCTGATGAACAAA 634  
Db 598 ttgttcagagaatcgtaaaccagctacttagtttacaatcagggagaatttaagatacca 657  
Qy 635 GTGATATATAGCATCTTGATGATGTTGGGACACACAGGAATTTGTCAAAAACCTGTGCAA 694  
Db 658 gcgtccaaagacatcgcagctggggatcatcgcagaatttggaaatcgtctaccaca 717  
Qy 695 AGTCTATAGTAAATGGGGATCTTGTGTGGGCAACTCCGAGCGGCAACGACATACCT 754  
Db 718 aatcactagtaaatgagtgcttcttgcctcaaggcgaacacgcgaataatgccatacct 777  
Qy 755 TCAATTTGCGATCGCTGGTAAACCTTTACCTGATTCAGAAAGAGAGGCTCATATACC 814  
Db 778 ttaactgcaagatgcttaaltcaccc-----tccagtagagccaggccgagagac 828  
Qy 815 AGGAAGCTCATCAACAATATGAACATATGACGTCTTCGCTCTCACAACCAAGTCCA 874  
Db 829 aagaagcttgcacgctgataagtagtaagtagtcttcaactgtgtcagcagcaataca 888  
Qy 875 TCAAGAAGAGAGAGAAAGATTTGCGATCTGCTTGAATTTGGCGGCAAGAAAGTTCCCA 934  
Db 889 ttcaagagatgagaagaatttccaagtcagtctcgatttgcagcagcgatlaacctc 948  
Qy 935 TGAAGAAAGACAGTCTTCCCTCATTCAGAAAGTTTATCTACCTGCCAGATCTCCAG 994  
Db 949 g-----cctcagcatltaggggtgtagaactcctttagcacaagcagatactacag 1002  
Qy 995 GCAAGATCACCTCTCTGATTCACAGCACATGACAGCAGCCATTAACACAGCTGGAGG 1054  
Db 1003 gtaaatcatctcatctatctactagttccctggaagctgctggaagaactgttgggaag 1062  
Qy 1055 ACCTGTTAAGAAAGTGTTCAGAAAGTTTCATGCGCAGCAAGAGAAATCTGTGCTT 1114  
Db 1063 attagtagagaatgcatcttacttcttccaacctcagggaagaac-----acctt 1119  
Qy 1115 ATGCTTAAGAGGCATCATCATGATGATGATGACAGACAGATTGGCATTCAGTCAATCTATC 1174  
Db 1120 atgccagacagctgttccaagaatgtagtactcgtggcactgcctccacgccccctcata 1179  
Qy 1175 GTTTTCTCTGTGATGAGCGACTCTTGTGCTGCAACAACGAAGCAAACTCATCCGTT 1234  
Db 1180 gattcataatgaaatgagtaggaacatgcttagcgccacacaagaatgtaaaacttgcacc 1239  
Qy 1235 CTCAGACTACTAATGAACCTCAACTGTGAATATATTTTACATATCTCTTCACAGAGAGAGA 1294





Db 1151 99 1152

RESULT 9

AK84490

ID AK84490 standard; DNA; 3806 BP.

XX

AC AK84490;

XX

DT 07-NOV-2001 (first entry)

XX

DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:39302.

XX

KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;

KW cytostatic; gene therapy; vaccine; metastasis; ds.

XX

OS Homo sapiens.

XX

PN W0200157182-A2.

XX

PD 09-AUG-2001.

XX

PF 17-JAN-2001; 2001WO-US01354.

XX

PR 31-JAN-2000; 2000US-0179065.

PR 04-FEB-2000; 2000US-0180628.

PR 24-FEB-2000; 2000US-0184664.

PR 02-MAR-2000; 2000US-0186350.

PR 16-MAR-2000; 2000US-0189874.

PR 17-MAR-2000; 2000US-0190076.

PR 18-APR-2000; 2000US-0198123.

PR 19-MAY-2000; 2000US-0205515.

PR 07-JUN-2000; 2000US-0209467.

PR 28-JUN-2000; 2000US-0214886.

PR 30-JUN-2000; 2000US-0215135.

PR 07-JUL-2000; 2000US-0216647.

PR 07-JUL-2000; 2000US-0216880.

PR 11-JUL-2000; 2000US-0217487.

PR 11-JUL-2000; 2000US-0217496.

PR 14-JUL-2000; 2000US-0218280.

PR 26-JUL-2000; 2000US-0220963.

PR 26-JUL-2000; 2000US-0220964.

PR 14-AUG-2000; 2000US-0224518.

PR 14-AUG-2000; 2000US-0224519.

PR 14-AUG-2000; 2000US-0225213.

PR 14-AUG-2000; 2000US-0225214.

PR 14-AUG-2000; 2000US-0225266.

PR 14-AUG-2000; 2000US-0225267.

PR 14-AUG-2000; 2000US-0225268.

PR 14-AUG-2000; 2000US-0225270.

PR 14-AUG-2000; 2000US-0225447.

PR 14-AUG-2000; 2000US-0225757.

PR 14-AUG-2000; 2000US-0225758.

PR 14-AUG-2000; 2000US-0225759.

PR 18-AUG-2000; 2000US-0226279.

PR 22-AUG-2000; 2000US-0226681.

PR 23-AUG-2000; 2000US-0226686.

PR 23-AUG-2000; 2000US-0227182.

PR 23-AUG-2000; 2000US-0227009.

PR 30-AUG-2000; 2000US-0228924.

PR 01-SEP-2000; 2000US-0229287.

PR 01-SEP-2000; 2000US-0229343.

PR 01-SEP-2000; 2000US-0229344.

PR 01-SEP-2000; 2000US-0229345.

PR 05-SEP-2000; 2000US-0229509.

PR 05-SEP-2000; 2000US-0229513.

PR 06-SEP-2000; 2000US-0230437.

PR 06-SEP-2000; 2000US-0230438.

PR 08-SEP-2000; 2000US-0231242.

PR 08-SEP-2000; 2000US-0231243.

PR 08-SEP-2000; 2000US-0231244.

PR 08-SEP-2000; 2000US-0231413.

PR 08-SEP-2000; 2000US-0231414.

PR 08-SEP-2000; 2000US-0232080.

PR 08-SEP-2000; 2000US-0232081.

PR 12-SEP-2000; 2000US-0231968.

PR 14-SEP-2000; 2000US-0232397.

PR 14-SEP-2000; 2000US-0232398.

PR 14-SEP-2000; 2000US-0232399.

PR 14-SEP-2000; 2000US-0232400.

PR 14-SEP-2000; 2000US-0232401.

PR 14-SEP-2000; 2000US-0233063.

PR 14-SEP-2000; 2000US-0233064.

PR 14-SEP-2000; 2000US-0233065.

PR 21-SEP-2000; 2000US-0234223.

PR 21-SEP-2000; 2000US-0234274.

PR 25-SEP-2000; 2000US-0234997.

PR 25-SEP-2000; 2000US-0234998.

PR 26-SEP-2000; 2000US-0235484.

PR 27-SEP-2000; 2000US-0235834.

PR 27-SEP-2000; 2000US-0235836.

PR 29-SEP-2000; 2000US-0236327.

PR 29-SEP-2000; 2000US-0236367.

PR 29-SEP-2000; 2000US-0236368.

PR 29-SEP-2000; 2000US-0236369.

PR 29-SEP-2000; 2000US-0236370.

PR 02-OCT-2000; 2000US-0236802.

PR 02-OCT-2000; 2000US-0237037.

PR 02-OCT-2000; 2000US-0237038.

PR 02-OCT-2000; 2000US-0237039.

PR 02-OCT-2000; 2000US-0237040.

PR 13-OCT-2000; 2000US-0239935.

PR 13-OCT-2000; 2000US-0239937.

PR 20-OCT-2000; 2000US-0240960.

PR 20-OCT-2000; 2000US-0241221.

PR 20-OCT-2000; 2000US-0241785.

PR 20-OCT-2000; 2000US-0241786.

PR 20-OCT-2000; 2000US-0241787.

PR 20-OCT-2000; 2000US-0241808.

PR 20-OCT-2000; 2000US-0241809.

PR 20-OCT-2000; 2000US-0241826.

PR 01-NOV-2000; 2000US-0244617.

PR 08-NOV-2000; 2000US-0246474.

PR 08-NOV-2000; 2000US-0246475.

PR 08-NOV-2000; 2000US-0246476.

PR 08-NOV-2000; 2000US-0246477.

PR 08-NOV-2000; 2000US-0246478.

PR 08-NOV-2000; 2000US-0246523.

PR 08-NOV-2000; 2000US-0246524.

PR 08-NOV-2000; 2000US-0246525.

PR 08-NOV-2000; 2000US-0246526.

PR 08-NOV-2000; 2000US-0246527.

PR 08-NOV-2000; 2000US-0246528.

PR 08-NOV-2000; 2000US-0246532.

PR 08-NOV-2000; 2000US-0246609.

PR 08-NOV-2000; 2000US-0246610.

PR 08-NOV-2000; 2000US-0246611.

PR 08-NOV-2000; 2000US-0246613.

PR 17-NOV-2000; 2000US-0249207.

PR 17-NOV-2000; 2000US-0249208.

PR 17-NOV-2000; 2000US-0249209.

PR 17-NOV-2000; 2000US-0249210.

PR 17-NOV-2000; 2000US-0249211.

PR 17-NOV-2000; 2000US-0249212.

PR 17-NOV-2000; 2000US-0249213.

PR 17-NOV-2000; 2000US-0249214.

PR 17-NOV-2000; 2000US-0249215.

PR 17-NOV-2000; 2000US-0249216.

PR 17-NOV-2000; 2000US-0249217.

PR 17-NOV-2000; 2000US-0249218.

PR 17-NOV-2000; 2000US-0249244.

PR 17-NOV-2000; 2000US-0249245.

PR 17-NOV-2000; 2000US-0249264.

PR 17-NOV-2000; 2000US-0249265.

PR 17-NOV-2000; 2000US-0249297.

PR 17-NOV-2000; 2000US-0249299.

PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250160.  
PR 05-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Barash SC, Ruben SM;  
XX  
XX WPI; 2001-483426/52.  
XX  
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
XX useful for preventing, diagnosing and/or treating cancers and  
XX metastasis -  
XX  
XX  
XX Disclosure; SEQ ID NO 39302; 3071pp + Sequence Listing; English.  
XX  
XX AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)  
XX amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic  
XX activity, and can be used in gene therapy and vaccine production. (I)  
XX proteins and polynucleotides may be used in the prevention, diagnosis and  
XX treatment of diseases associated with inappropriate (I) expression. For  
XX example, they may be used to treat disorders associated with decreased  
XX expression by rectifying mutations or deletions in a patient's genome  
XX that affect the activity of (I) by expressing inactive proteins or to  
XX supplement the patients own production of (I). Additionally, (I)  
XX polynucleotides may be used to produce the secreted (I), by inserting  
XX the nucleic acids into a host cell and culturing the cell to express the  
XX protein. (I) proteins and polynucleotides may be used to prevent,  
XX diagnose and treat immune/haematopoietic-related diseases, especially  
XX cancers and cancer metastases of haematopoietic-derived cells. AAK64703  
XX to AAK87694 represent human immune/haematopoietic antigen genomic  
XX sequences from the present invention. AAK54942 to AAK54950 and AAM82169  
XX represent sequences used in the exemplification of the present invention.  
SQ Sequence 3806 BP; 1149 A; 672 C; 773 G; 1212 T; 0 other;

Query Match 3.1%; Score 189; DB 22; Length 3806;  
Best Local Similarity 100.0%; Pred. No. 2.4e-41;  
Matches 189; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 702 AGTAATGGGGGATCTTGCTGGCGAAGCTCGAGCGGAGACAGCATACCTTCAATTG 761  
DB 3618 agtaaatgggggattcttgctggcgaaacctcgagcggaacacgacatccctcaatly 3677  
QY 762 TCCGATGCTGTGTAACCTTACCTGATTCAGAGAGAGAGGTCATGATTAACGAGAAC 821  
DB 3678 tcgtagctgtgtaaaccttacctgattcagaagaggaaggtcattacacaggaagc 3737  
QY 822 TCATCGAATAATTAACACTTATGTCAGTGTCTGCTCTCAACCAAAAGTCATCAAGA 881  
DB 3738 tcctcgaaatatgaactatgtagctgtctgtctcacaacaagatccatcaaga 3797  
QY 882 AGAAGGAGA 890  
DB 3798 agaagagaga 3806

RESULT 10  
AAK84489  
ID AAK84489 standard; DNA; 3815 BP.  
XX

AC AAK84489;  
XX  
XX 07-NOV-2001 (first entry)  
XX  
XX Human immune/haematopoietic antigen genomic sequence SEQ ID NO:39301.  
DE  
XX  
XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;  
KW cytostatic; gene therapy; vaccine; metastasis; ds.  
XX  
XX Homo sapiens.  
XX  
XX WO200157182-A2.  
XX  
XX 09-AUG-2001.  
XX  
XX  
XX 17-JAN-2001; 2001WO-US01354.  
XX  
XX 31-JAN-2000; 2000US-0179065.  
XX 04-FEB-2000; 2000US-0180628.  
XX 24-FEB-2000; 2000US-0184664.  
XX 02-MAR-2000; 2000US-0186350.  
XX 16-MAR-2000; 2000US-0189874.  
XX 17-MAR-2000; 2000US-0190076.  
XX 18-APR-2000; 2000US-0198123.  
XX 19-MAY-2000; 2000US-0205515.  
XX 07-JUN-2000; 2000US-0209467.  
XX 28-JUN-2000; 2000US-0214886.  
XX 30-JUN-2000; 2000US-0215135.  
XX 07-JUL-2000; 2000US-0216647.  
XX 07-JUL-2000; 2000US-0216880.  
XX 11-JUL-2000; 2000US-0217487.  
XX 11-JUL-2000; 2000US-0217496.  
XX 14-JUL-2000; 2000US-0218290.  
XX 26-JUL-2000; 2000US-0220963.  
XX 26-JUL-2000; 2000US-0220964.  
XX 14-AUG-2000; 2000US-0224518.  
XX 14-AUG-2000; 2000US-0224519.  
XX 14-AUG-2000; 2000US-0225213.  
XX 14-AUG-2000; 2000US-0225214.  
XX 14-AUG-2000; 2000US-0225266.  
XX 14-AUG-2000; 2000US-0225267.  
XX 14-AUG-2000; 2000US-0225268.  
XX 14-AUG-2000; 2000US-0225270.  
XX 14-AUG-2000; 2000US-0225447.  
XX 14-AUG-2000; 2000US-0225757.  
XX 14-AUG-2000; 2000US-0225758.  
XX 14-AUG-2000; 2000US-0225759.  
XX 18-AUG-2000; 2000US-0226279.  
XX 22-AUG-2000; 2000US-0226881.  
XX 22-AUG-2000; 2000US-0226882.  
XX 22-AUG-2000; 2000US-0227182.  
XX 23-AUG-2000; 2000US-0227009.  
XX 30-AUG-2000; 2000US-0228924.  
XX 01-SEP-2000; 2000US-0229287.  
XX 01-SEP-2000; 2000US-0229343.  
XX 01-SEP-2000; 2000US-0229344.  
XX 01-SEP-2000; 2000US-0229345.  
XX 05-SEP-2000; 2000US-0229509.  
XX 05-SEP-2000; 2000US-0229513.  
XX 06-SEP-2000; 2000US-0230437.  
XX 06-SEP-2000; 2000US-0230438.  
XX 08-SEP-2000; 2000US-0231242.  
XX 08-SEP-2000; 2000US-0231243.  
XX 08-SEP-2000; 2000US-0231244.  
XX 08-SEP-2000; 2000US-0231413.  
XX 08-SEP-2000; 2000US-0231414.  
XX 08-SEP-2000; 2000US-0232080.  
XX 08-SEP-2000; 2000US-0232081.  
XX 12-SEP-2000; 2000US-0231968.  
XX 14-SEP-2000; 2000US-0232397.  
XX 14-SEP-2000; 2000US-0232398.  
XX 14-SEP-2000; 2000US-0232399.  
XX 14-SEP-2000; 2000US-0232400.

PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249267.  
PR 17-NOV-2000; 2000US-0249269.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251968.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.

PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251988.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Barash SC, Ruben SM;  
XX  
XX WPI; 2001-483426/52.  
XX  
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
XX useful for preventing, diagnosing and/or treating cancers and  
XX metastasis -  
XX  
XX Disclosure; SEQ ID NO 39301; 3071bp + Sequence Listing; English.  
XX  
XX  
XX AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)  
XX amino acid sequences given in AAM82170 to AAM91921. (I) have cytosolic  
XX activity, and can be used in gene therapy and vaccine production. (I)  
XX proteins and polynucleotides may be used in the prevention, diagnosis and  
XX treatment of diseases associated with inappropriate (I) expression. For  
XX example, they may be used to treat disorders associated with decreased  
XX expression by rectifying mutations or deletions in a patient's genome  
XX that affect the activity of (I) by expressing inactive proteins or to  
XX supplement the patients own production of (I). Additionally, (I)  
XX polynucleotides may be used to produce the secreted (I), by inserting  
XX the nucleic acids into a host cell and culturing the cell to express the  
XX protein. (I) proteins and polynucleotides may be used to prevent,  
XX diagnose and treat immune/hematopoietic-related diseases, especially  
XX cancers and cancer metastases of haematopoietic-derived cells. AAK64703  
XX to AAK87694 represent human immune/hematopoietic antigen genomic  
XX sequences from the present invention. AAK54942 to AAK54950 and AAM82169  
XX represent sequences used in the exemplification of the present invention.  
XX  
XX Sequence 3815 BP; 1141 A; 654 C; 785 G; 1235 T; 0 other;

Query Match 3.0%; Score 187.4; DB 22; Length 3815;  
Best Local Similarity 99.5%; Pred. No. 6.6e-41;  
Matches 188; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 702 AGTAAATGGGGATCTGTGCTGGCAACTCCGAGCGGAGACCACTTACCTTCAATG 761  
Db 3627 agtaaatggggatctgtgctggcaacctccgagcggaacagccatacctcaatg 3686  
QY 762 TCGATGCTGTTAAACCTTTACCTGATTCAGAGAGAGGGTCAATGTAACACAGAGAC 821  
Db 3687 tcgatactgtglaaaaccttaacctgattcagaagaagaggtctgtataccaggaagc 3746  
QY 822 TCATCAGAAATATGAACTATGACAGTCTGCTCTCTCAACCAAGTCCATCAAGA 881  
Db 3747 tcattcagaataatgaaactatgacagtgcttgcgtctctcaaccaaagtcattcaaga 3806  
QY 882 AGAAGGAGA 890  
Db 3807 agaagagaga 3815

RESULT 11  
AAH04011  
ID AAH04011 standard; cDNA; 767 BP.  
XX  
XX AAH04011;  
AC  
XX  
XX 26-JUN-2001 (first entry)  
DT  
XX  
XX Human cDNA clone (5'-primer) SEQ ID NO:846.  
DE  
XX  
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

```
XX OS Homo sapiens.
XX PN EPI074617-A2.
XX PD 07-FEB-2001.
XX PF 28-JUL-2000; 2000EP-0116126.
XX PR 29-JUL-1999; 99JP-0248036.
XX PR 27-AUG-1999; 99JP-0300253.
XX PR 11-JAN-2000; 2000JP-0118776.
XX PR 02-MAY-2000; 2000JP-0183767.
XX PR 09-JUN-2000; 2000JP-0241899.
XX PA (HELI-) HELIX RES INST.
XX PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX PD WPI; 2001-318749/34.
XX PF 28-JUL-2000; 2000EP-0116126.
XX PR 29-JUL-1999; 99JP-0248036.
XX PR 27-AUG-1999; 99JP-0300253.
XX PR 11-JAN-2000; 2000JP-0118776.
XX PR 02-MAY-2000; 2000JP-0183767.
XX PR 09-JUN-2000; 2000JP-0241899.
XX PS Claim 1; SEQ ID 846; 2537pp + CD ROM; English.
XX CC The present invention describes primer sets for synthesizing 5602
XX CC full-length cDNAs defined in the specification. Where a primer set
XX CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
XX CC to the complementary strand of a polynucleotide which comprises one of
XX CC the 5602 nucleotide sequences defined in the specification, where the
XX CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
XX CC of an oligonucleotide comprising a sequence complementary to the
XX CC complementary strand of a polynucleotide which comprises a 5'-end
XX CC sequence and an oligonucleotide comprising a sequence complementary to a
XX CC polynucleotide which comprises a 3'-end sequence, where the
XX CC oligonucleotide comprises at least 15 nucleotides and the combination of
XX CC the 5'-end sequence/3'-end sequence is selected from those defined in
XX CC the specification. The primer sets can be used in antisense therapy and
XX CC in gene therapy. The primers are useful for synthesizing polynucleotides,
XX CC particularly full-length cDNAs. The primers are also useful for the
XX CC detection and/or diagnosis of the abnormality of the proteins encoded by
XX CC the full-length cDNAs. The primers allow obtaining of the full-length
XX CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
XX CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
XX CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
XX CC represent oligonucleotides, all of which are used in the exemplification
XX CC of the present invention.
XX SQ Sequence 767 BP; 242 A; 151 C; 179 G; 192 T; 3 other;
XX
XX Query Match 2.4%; Score 150.6; DB 22; Length 767;
XX Best Local Similarity 89.5%; Pred. No. 4.1e-31;
XX Matches 162; Conservative 0; Mismatches 19; Indels 0; Gaps 0;
```

```
DB 556 g 556
XX RESULT 12
XX AAH17184
XX ID AAH17184 standard; cDNA; 1367 BP.
XX AC AAH17184;
XX DT 26-JUN-2001 (first entry)
XX DE Human CDNA sequence SEQ ID NO:16546.
XX KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX OS Homo sapiens.
XX PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX PD WPI; 2001-318749/34.
XX PF 28-JUL-2000; 2000EP-0116126.
XX PR 29-JUL-1999; 99JP-0248036.
XX PR 27-AUG-1999; 99JP-0300253.
XX PR 11-JAN-2000; 2000JP-0118776.
XX PR 02-MAY-2000; 2000JP-0183767.
XX PR 09-JUN-2000; 2000JP-0241899.
XX PS (HELI-) HELIX RES INST.
XX CC The present invention describes primer sets for synthesizing 5602
XX CC full-length cDNAs defined in the specification. Where a primer set
XX CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
XX CC to the complementary strand of a polynucleotide which comprises one of
XX CC the 5602 nucleotide sequences defined in the specification, where the
XX CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
XX CC of an oligonucleotide comprising a sequence complementary to the
XX CC complementary strand of a polynucleotide which comprises a 5'-end
XX CC sequence and an oligonucleotide comprising a sequence complementary to a
XX CC polynucleotide which comprises a 3'-end sequence, where the
XX CC oligonucleotide comprises at least 15 nucleotides and the combination of
XX CC the 5'-end sequence/3'-end sequence is selected from those defined in
XX CC the specification. The primer sets can be used in antisense therapy and
XX CC in gene therapy. The primers are useful for synthesizing polynucleotides,
XX CC particularly full-length cDNAs. The primers are also useful for the
XX CC detection and/or diagnosis of the abnormality of the proteins encoded by
XX CC the full-length cDNAs. The primers allow obtaining of the full-length
XX CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
XX CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
XX CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
XX CC represent oligonucleotides, all of which are used in the exemplification
XX CC of the present invention.
XX SQ Sequence 1367 BP; 411 A; 280 C; 326 G; 350 T; 0 other;
XX
XX Query Match 2.4%; Score 150.6; DB 22; Length 1367;
XX Best Local Similarity 89.5%; Pred. No. 6.1e-31;
XX Matches 162; Conservative 0; Mismatches 19; Indels 0; Gaps 0;
```

QY	958	TCATCAGAAAGTTTACTCTGCGCAGAGATCTCCAAAGGCAAGATCAGCTCTGGATACC	1017		
Db	376	lcaeaatctgaagtataaetctgctgttactactaataagcaagatcaacgctctctgataacc	435		
QY	1018	AGCACCATGAGAGAGGCGCATGAACCGAGGCGGGAGAGCCGGTAAAGGTGATTCAG	1077		
Db	436	agcacatgagagagagcgcatagaaccagagctcggaaggacccggtaaagaagtgattcag	495		
QY	1078	AAGTTCATGGCGAGCATGAGAGGAAATCTGTCCTCATGCTTAAGAGAGCATCATGAAA	1137		
Db	466	aagttccatgtgcgcagcatgaagggaatctgtgctccatgctaaagagcatcatcagaa	555		
QY	1138	G 1138			
Db	556	g 556			
RESULT 13					
ID	AAH99288	AAH99288 standard; cDNA; 435 BP.			
XX	AAH99288;				
DT	16-OCT-2001	(first entry)			
DE	Human protein encoding cDNA sequence SEQ ID NO:123.				
XX					
KW	Human; cancer; ulcer; HIV infection; human immunodeficiency virus; antinflmmatory; antirheumatic; antiarthritic; immunosuppressive; antibacterial; endocrine; cardiac; central nervous system; virucide; anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; angiaregant; haemostatic; vulnary; antitumor; osteopathic; eczema; dermatological; antiallergic; antidiabetic; antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic; antiparkinsonian; infection immunostimulant; gene therapy; antisense therapy; vaccine; inflammation; antiapophytic; rheumatoid arthritis; septic shock; pancreatitis; cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity; genetic disease; haematopoietic disorder; platelet disorder; asthma; thrombocytopenia; osteoporosis; severe combined immunodeficiency; allergic rhinitis; diabetes; multiple sclerosis; depression; Alzheimer's disease; Parkinson's disease; neurodegenerative disorder; neurological disorder; ss.				
OS	Homo sapiens.				
XX					
PN	WO200153455-A2.				
PD	26-JUL-2001.				
XX					
PF	22-DEC-2000; 2000MO-US35017.				
XX					
PR	23-DEC-1999; 99US-0471275.				
PR	21-JAN-2000; 2000US-0488725.				
PR	25-APR-2000; 2000US-0552317.				
XX					
PA	(HYSE-) HYSEQ INC.				
XX					
PI	Tang YT, Liu C, Drmanac RT;				
XX					
DR	WPI: 2001-457603/49.				
XX					
DR	P-PSDB: AAM25347.				
XX					
PT	Isolated human polynucleotides encoding polypeptides, useful for the treatment and diagnosis of e.g. cancer, ulcers and HIV infection -				
PS	Claim 1; Page 354; 1217pp; English.				
XX					
XX	AAH99166 to AAH99904 encode the human proteins given in AAM25225 to AAM25663. The proteins can have activities based on the tissues and cells they are expressed in, such as: antinflmmatory; antirheumatic; antiarthritic; immunosuppressive; antibacterial; endocrine; cardiac; central nervous system; virucide; anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; angiaregant; haemostatic; vulnary;				

CC	antitumor; osteopathic; dermatological; antiallergic; antiasthmatic;
CC	antidiabetic; cytosolic; neuroprotective; antidepressant; nootropic;
CC	antiparkinsonian; and immunostimulant. The proteins and polynucleotides
CC	encoding them can be used in gene therapy, antisense therapy and vaccine
CC	production. The proteins and polynucleotides are useful for screening for
CC	agonists or antagonists of a protein and for the treatment and diagnosis
CC	of disorders associated with the activity of a protein e.g. Inflammation,
CC	rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,
CC	neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal
CC	infections, autoimmunity, genetic diseases, haematopoietic disorders,
CC	anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers,
CC	osteoporosis, severe combined immunodeficiency, eczema, allergic
CC	rheinitis, asthma, diabetes, cancer, multiple sclerosis, depression,
CC	Alzheimer's disease, Parkinson's disease, neurodegenerative and
CC	neurological disorders.
XX	
SO	Sequence 435 BP; 126 A; 70 C; 92 G; 147 T; 0 other;
	Query Match 1.7%; Score 106; DB 22; Length 435;
	Best Local Similarity 100.0%; Pred. No. 7.9e-19;
	Matches 106; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY	421 GAGAAAGCAGCAGCTGCCACATAGATGAAGTGCAGAAAGTCAGATGTATCTCTACAGGG 480
Db	1 gagaagaagcagcagctgcacacatagatgaagtcagaagaagtcagatcctcctcaag 60
OY	481 CAGGTCATCATGACAGAGATGCGTGGGGCCTTAGTGTCTTGAGG 526
Db	61 cagggtgcatcagcaagatgcgtggtggtcctatgatgtcttgaag 106
	RESULT 14
ID	AAx80993
AC	AAx80993 standard; DNA: 3361 BP.
XX	
XX	AAx80993:
DT	06-SEP-1999 (first entry)
XX	
XX	Mouse steroid receptor coactivator-3 (SRC-3) partial DNA sequence.
DE	
XX	Steroid receptor coactivator-3; SRC-3; transcription enhancement; human;
KM	hormone response element; estrogen receptor alpha; neoplastic state;
KW	mouse; ss.
KX	
OS	Mus sp.
XX	
PN	W09332621-A2.
XX	
PD	01-JUL-1999.
XX	
XX	01-DEC-1998; 98WO-US25478.
PF	
PR	22-DEC-1997; 97US-0068511.
XX	
PA	(AMHP ) AMERICAN HOME PROD CORP.
XX	
PI	Frail DE, Lytle CR, Suen C;
XX	
DR	WPI: 1999-405171/34.
XX	
PT	Nucleic acid encoding a steroid receptor co-activator-3, useful for
PC	determining the neoplastic states of cells in humans or animals
XX	
PS	Example 9; Page 73-75; 75pp; English.
XX	
CC	The invention relates to a human steroid receptor coactivator-3 (SRC-3).
CC	Host cells transformed with vectors comprising the SRC-3 gene can be used
CC	for the recombinant production of the SRC-3 protein. SRC-3 can be used to
CC	enhance the transcription of genes, which are regulated by hormone
CC	response elements, e.g. estrogen receptor alpha. SRC-3 can be used in
CC	methods to screen for compounds that bind to SRC-3 or which modulate





XX 30-JAN-2001; 2001WO-US00667.  
 XX  
 PF 04-FEB-2000; 2000US-0180312.  
 XX  
 PR 26-MAY-2000; 2000US-0207456.  
 XX  
 PR 30-JUN-2000; 2000US-0608408.  
 XX  
 PR 03-AUG-2000; 2000US-0632366.  
 XX  
 PR 21-SEP-2000; 2000US-0234687.  
 XX  
 PR 27-SEP-2000; 2000US-0236359.  
 XX  
 PR 04-OCT-2000; 2000GB-0024263.  
 XX  
 PA (MOLE-) MOLECULAR DYNAMICS INC.  
 XX  
 PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX  
 DR WPI; 2001-483446/52.  
 XX  
 PT Single exon nucleic acid probes for analyzing gene expression in human  
 XX  
 PT brains -  
 XX  
 PS Example 4; SEQ ID NO: 16870; 650bp + Sequence Listing; English.  
 XX  
 CC The present invention provides a number of single exon nucleic acid  
 CC probes which are derived from genomic sequences expressed in the human  
 CC brain. They can be used to measure gene expression in brain cell samples,  
 CC which may enable the diagnosis and improved treatment of nervous system  
 CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,  
 CC epilepsy and cancers. The present sequence is one of the probes of the  
 CC invention.  
 XX  
 XX Sequence 330 BP; 81 A; 80 C; 73 G; 96 T; 0 other;  
 SQ

Query Match 1.1%; Score 70.6; DB 22; Length 330;  
 Best Local Similarity 62.2%; Pred. No. 5e-09;  
 Matches 150; Conservative 0; Mismatches 79; Indels 12; Gaps 2;  
 QY 2193 TACACATGGAACCTCGCTCAAGAGAGCATAAATTGTCACAGACTCTTGACAGACAG 2252  
 DB 242 TATGATGAGGTCTACTGTTCACAGAGAGCACCAGATTGTCACAAAGTTGCTGCAGAAATGG 183  
 QY 2253 CAGTCCCTGCTGGACTTGCCCAAGTTAACAGAGAGCCACAGGCAAAACCTGAGCCCA 2312  
 DB 182 GAATTTCCACGCTGAGGTAGCCAGATTACTGCAAGAGCCACGAGGAAAGACACCCAG-- 124  
 QY 2313 GGAGTCCAGCAGCAGCAGCTCTGATCAGAGTACTATTAAACAAAGCCGCGTAGACCC 2372  
 DB 125 -----CAGTATTAACCTTCTGTGGGAGCGAATGTTGTCAAGCAGAGCAGCTAGTCC 72  
 QY 2373 CAGAGAGAAAGAG--AATGCACTACTTCGCTATTGCTAGATTAAGATGATCTAAAGA 2429  
 DB 71 TAGAGAGAGAGAGAAATATGATCTTCTAGATACCTGCTGACAGAGGATGATCTAGTGA 12  
 QY 2430 T 2430  
 DB 11 T 11

RESULT 19  
 AAK42647/C  
 ID AAK42647 standard; DNA; 330 BP.  
 XX  
 AC AAK42647;  
 XX  
 DT 06-NOV-2001 (first entry)  
 XX  
 DE Human bone marrow expressed single exon probe SEQ ID NO: 17204.  
 XX  
 KW Human; bone marrow expressed exon; gene expression analysis; probe;  
 XX  
 OS microarray; cancer; leukaemia; lymphoma; myeloma; ss.  
 XX  
 XX Homo sapiens.  
 XX

PN W0200157276-A2.  
 XX  
 PD 09-AUG-2001.  
 XX  
 XX 30-JAN-2001; 2001WO-US00668.  
 PF  
 XX  
 PR 04-FEB-2000; 2000US-0180312.  
 XX  
 PR 26-MAY-2000; 2000US-0207456.  
 XX  
 PR 30-JUN-2000; 2000US-0608408.  
 XX  
 PR 03-AUG-2000; 2000US-0632366.  
 XX  
 PR 21-SEP-2000; 2000US-0234687.  
 XX  
 PR 27-SEP-2000; 2000US-0236359.  
 XX  
 PR 04-OCT-2000; 2000GB-0024263.  
 XX  
 PA (MOLE-) MOLECULAR DYNAMICS INC.  
 XX  
 PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX  
 DR WPI; 2001-488900/53.  
 XX  
 PT Human genome-derived single exon nucleic acid probes useful for  
 XX  
 PT analyzing gene expression in human bone marrow -  
 XX  
 PS Example 4; SEQ ID NO: 17204; 658bp + Sequence Listing; English.  
 XX  
 CC The present invention provides a number of single exon nucleic acid  
 CC probes which are derived from genomic sequences expressed in the human  
 CC bone marrow. They can be used to measure gene expression in bone marrow  
 CC samples, which may enable the improved diagnosis and treatment of cancers  
 CC such as lymphoma, leukaemia and myeloma. The present sequence is one of  
 CC the probes of the invention.  
 XX  
 XX Sequence 330 BP; 81 A; 80 C; 73 G; 96 T; 0 other;  
 SQ

Query Match 1.1%; Score 70.6; DB 22; Length 330;  
 Best Local Similarity 62.2%; Pred. No. 5e-09;  
 Matches 150; Conservative 0; Mismatches 79; Indels 12; Gaps 2;  
 QY 2193 TACACATGGAACCTCGCTCAAGAGAGCATAAATTGTCACAGACTCTTGACAGACAG 2252  
 DB 242 TATGATGAGGTCTACTGTTCACAGAGAGCACCAGATTGTCACAAAGTTGCTGCAGAAATGG 183  
 QY 2253 CAGTCCCTGCTGGACTTGCCCAAGTTAACAGAGAGCCACAGGCAAAACCTGAGCCCA 2312  
 DB 182 GAATTTCCACGCTGAGGTAGCCAGATTACTGCAAGAGCCACGAGGAAAGACACCCAG-- 124  
 QY 2313 GGAGTCCAGCAGCAGCAGCTCTGATCAGAGTACTATTAAACAAAGCCGCGTAGACCC 2372  
 DB 125 -----CAGTATTAACCTTCTGTGGGAGCGAATGTTGTCAAGCAGAGCAGCTAGTCC 72  
 QY 2373 CAGAGAGAAAGAG--AATGCACTACTTCGCTATTGCTAGATTAAGATGATCTAAAGA 2429  
 DB 71 TAGAGAGAGAGAGAAATATGATCTTCTAGATACCTGCTGACAGAGGATGATCTAGTGA 12  
 QY 2430 T 2430  
 DB 11 T 11

RESULT 20  
 AA123403/C  
 ID AA123403 standard; DNA; 330 BP.  
 XX  
 AC AA123403;  
 XX  
 DT 12-OCT-2001 (first entry)  
 XX  
 DE Probe #13336 for gene expression analysis in human cervical cell sample.  
 XX  
 KW Probe; human; microarray; gene expression; cervical epithelial cell;  
 XX  
 OS cervical cancer; ss.  
 XX



```
OS Homo sapiens.
XX
XX MO200157278-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00670.
XX
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488901/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human cervical epithelial cells -
XX
XX Claim 25; SEQ ID No 13336; 487bp; English.
XX
XX The present invention relates to human single exon nucleic acid probes
XX (SENPs). The present sequence is one such probe. The SENPs are derived
XX from human HeLa cells. The SENPs can be used to produce a single exon
XX microarray, which can be used for measuring human gene expression in a
XX sample derived from human cervical epithelial cells. By measuring gene
XX expression, the probes are therefore useful in grading and/or staging
XX of diseases of the cervix, notably cervical cancer.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at http://wipo.int/pub/published\_pcl\_sequences.
XX
XX Sequence 330 BP; 81 A; 80 C; 73 G; 96 T; 0 other;
XX
Query Match 1.1%; Score 70.6; DB 22; Length 330;
Best Local Similarity 62.2%; Pred. No. 5e-09;
Matches 150; Conservative 0; Mismatches 79; Indels 12; Gaps 2;
QY 2193 TACACATGGAACCTCGCTCAAGAGACATTAATAATTTGACACAGACTCTTGACAGACAG 2252
DB 242 TATGCATGGGTGCTACTGTTCACAGAGACACCGGATTTTGACAAAGTTGCTGCAGAAATGG 183
QY 2253 CAGTTCCCTGTGAGCTTGCCCAAGTTAAACAGAGAACCCACAGCAAAAGCTGAGCCA 2312
DB 182 GAATTCACCAAGCTGAGGAGCAAGATTACTGCAAGAACCCACTGGGAAAGACACCAG--- 124
QY 2313 GGAGTCCAGCAGACAGCTCCTCGATCAGAAAGTACTATTAAACAGAGACCGGTAGCCC 2372
DB 125 -----CAGTATTAACCTTCTGTGGGAGCAAGAAATGTTTCAAGCAGAGAGCTAAGTCC 72
QY 2373 CAAGAAGAAAGAG---AATGCACACTCTCGCTATTGCTAATAAAGTGTACTAAAGA 2429
DB 71 TAAGAAAGAGAGAAATATGACATCTTGTAGATCCTGCTGAGCAGGATGATCTTAAGTA 12
QY 2430 T 2430
DB 11 T 11
RESULT 21
AAI48724/c
ID AAI48724 standard; DNA: 330 BP.
XX
XX AAI48724;
XX
XX 17-OCT-2001 (first entry)
```

```
XX
XX Probe #17410 used to measure gene expression in human placenta sample.
DE
XX Probe; microarray; human; placenta; antenatal diagnosis;
XX genetic disorder; ss.
XX
XX Homo sapiens.
XX
XX MO200157272-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00663.
XX
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488897/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human placenta -
XX
XX Claim 25; SEQ ID No 17410; 654bp; English.
XX
XX The present invention relates to single exon nucleic acid probes (SENPs).
XX The present sequence is one such probe. The probes are useful for
XX producing a microarray for predicting, measuring and displaying gene
XX expression in samples derived from human placenta. The probes are useful
XX for antenatal diagnosis of human genetic disorders.
XX
XX Sequence 330 BP; 81 A; 80 C; 73 G; 96 T; 0 other;
XX
Query Match 1.1%; Score 70.6; DB 22; Length 330;
Best Local Similarity 62.2%; Pred. No. 5e-09;
Matches 150; Conservative 0; Mismatches 79; Indels 12; Gaps 2;
QY 2193 TACACATGGAACCTCGCTCAAGAGACATTAATAATTTGACACAGACTCTTGACAGACAG 2252
DB 242 TATGCATGGGTGCTACTGTTCACAGAGACACCGGATTTTGACAAAGTTGCTGCAGAAATGG 183
QY 2253 CAGTTCCCTGTGAGCTTGCCCAAGTTAAACAGAGAACCCACAGCAAAAGCTGAGCCA 2312
DB 182 GAATTCACCAAGCTGAGGAGCAAGATTACTGCAAGAACCCACTGGGAAAGACACCAG--- 124
QY 2313 GGAGTCCAGCAGACAGCTCCTCGATCAGAAAGTACTATTAAACAGAGACCGGTAGCCC 2372
DB 125 -----CAGTATTAACCTTCTGTGGGAGCAAGAAATGTTTCAAGCAGAGAGCTAAGTCC 72
QY 2373 CAAGAAGAAAGAG---AATGCACACTCTCGCTATTGCTAATAAAGTGTACTAAAGA 2429
DB 71 TAAGAAAGAGAGAAATATGACATCTTGTAGATCCTGCTGAGCAGGATGATCTTAAGTA 12
QY 2430 T 2430
DB 11 T 11
RESULT 22
AAI09030/c
ID AAI09030 standard; DNA: 330 BP.
XX
XX AAI09030;
XX
```











XX	AAV41256;	
AC		
XX		
DT	09-NOV-1998	(first entry)
XX		
DE	Human neuronal PAS domain protein NPAS1 cDNA.	
XX		
KW	NPAS1; neuronal bHLH-PAS domain; human; transcription factor;	
KW	therapy; diagnosis; ds.	
XX		
OS	Homo sapiens.	
XX		
FT	Key	Location/Qualifiers
FT	CDS	181..1953
FT		/*tag= a
XX		
PN	MO9831804-A1.	
XX		
PD	23-JUL-1998.	
XX		
PE	21-JAN-1998;	98MO-US01154.
XX		
PR	21-JAN-1997;	9705-0785310.
XX		
PA	(TEXA ) UNIV TEXAS SYSTEM.	
XX		
PI	McKnight SL, Russel DW;	
DR	WPI; 1998-414103/35.	
DR	P-PSDB; AAM68091.	
XX		
PT	New isolated neuronal PAS domain proteins - can regulate function of	
PT	neurological tissue such as brain tissue, used to develop products	
PT	for diagnosis and therapy	
XX		
PS	Claim 7; Page 21: 42pp; English.	
XX		
XX	This cDNA clone codes for human neuronal PAS domain protein NPAS1	
XX	(see AAM68091), a new member of the basic helix-loop-helix (bHLH)-PAS	
CC	family of transcription factors. A database search for expressed	
CC	sequence tags bearing sequence similarity to the PAS domain of the	
CC	aryl hydrocarbon receptor yielded 2 clones. These were used to	
CC	generate primers for PCR amplification of hybridisation probes,	
CC	and clones (see AAV41245-59) coding for human and mouse NPAS proteins	
CC	(see AAM68091-94) were isolated from mouse brain tissue, human brain	
CC	tissue and HeLa cells. The human NPAS1 gene was mapped to	
CC	chromosome 19q13.2-q13.3. The NPAS proteins can regulate the	
CC	function of neurological tissue such as brain tissue. They can be	
CC	produced recombinantly from transformed host cells or purified from	
CC	mammalian cells. NPAS proteins and polynucleotides can be used in	
CC	diagnosis (e.g. genetic hybridisation screens for NPAS transcripts),	
CC	therapy (e.g. gene therapy to modulate NPAS gene expression) and in	
CC	the biopharmaceutical industry (e.g. as immunogens, reagents for	
CC	isolating other transcriptional regulators, and reagents for	
CC	screening chemical libraries for lead pharmacological agents).	
XX		
XX	Sequence 2078 BP; 337 A; 751 C; 670 G; 320 T; 0 other;	

Query Match	0.98	Score 55.4	DB 19	Length 2078
Best Local Similarly	54.78	Pred. No. 0.00031		
Matches 110; Conservative	0	Mismatches 91	Indels 0	Gaps 0

[illegible]

Db	658	cacagctgagatgacgycgacagcgtcttcgaatacatcaccctcgggacactcag	757
Qy	671	AATTTGTCAAAACCTGCTGC	691
Db	758	aggtctcggagcaactcgggc	778

RESULT	31
AAZ34629	
ID	AAZ34629 standard; cDNA; 3486 BP.

AC	AAZ34629;
XX	
DT	15-FEB-2000 (first entry)
DD	

AA Human clock cDNA.  
DE  
XX  
KW CLOCK; transcription factor; circadian rhythm; human;  
KW jet lag; sleep disorder; depression; seasonal affective disorder  
KW fertility; therapy; ss.

OS	Homo sapiens.	
XX		
FH	Key	Location/Qualifiers
FT	CDS	358..2898

XX WO9957137-A1.  
PN  
XX  
PD 11-NOV-1999.

PF	06-MAY-1999;	99WO-US10072.
XX		
PR	07-MAY-1998;	98US-0084610.
XX		

XX Weitz CJ, Gekakis N, Staknis D;  
PI  
XX  
DR WPI; 2000-052938/04.

PT Novel heterodimeric composition for identifying modulators used in  
PT diagnosing and treating circadian clock disruption disorders -  
XX  
PS Disclosure; Fig 25A-B; 96pp; English.

CC This is the nucleotide sequence of human cDNA encoding CLOCK  
CC protein (see AA132215). CLOCK activates the transcription of the  
CC *per1* gene when present in combination with BMAL1 (see AA132209).  
CC The invention, based on the discovery of the transcriptional  
CC mechanism regulating genes responsible for the establishment and/or  
CC maintenance of the circadian clock, is useful for the assay of  
CC novel drugs aimed at restoration of a normal circadian cycle, the  
CC drugs being modulators of BMAL1-CLOCK-mediated transcription of  
CC E-box-linked genes. The drugs are used to treat conditions such as  
CC jet lag, sleep disorders, depression (seasonal affective disorder)  
CC and infertility. The invention also provides BMAL1 and CLOCK  
CC proteins with which to stimulate the transcription of an  
CC E-box-linked gene which regulates the circadian clock.

Sequence 3486 BP; 1097 A; 755 C; 721 G; 913 T; 0 other;

Query Match	0.98;	Score 55;	DB 21;	Length 3486;
Best Local Similarity	55.5%;			
Matches 106; Conservative	0;	Mismatches 85;	Indels 0;	Gaps 0;

496	AAGGATGCGCTGGGGCCATGATGCTTGAGAGGCCCTTGATGGGTTCTTCTTTGATGTAAC	555
Qy		
Db	aatgaaggttacacaatlaagttagagcctctcgtatggttttcttllttagcaatcatg	735
Qy	CTGGAAGCCAACTGTTGTGTTTGTCTCAGAGAATGTGACACAGATCTTAAGGTATACCAA	615
Qy	556	

Db 736 acagatggaagcataaataatgctcgcgagagcgttaactcattcaactgaacattacca 795  
 QY 616 GAAGAGCTGATGAACAAGATGATATAGCATCTTGGCATTTGGGACACACAGGAATTT 675  
 Db 796 tcgatctctgtgatacaaatatattatccaccagaagggaacattcaagagtc 855  
 QY 676 GTCAAAAACCT 686  
 Db 856 tataaataact 866  
 RESULT 32  
 AAV61450  
 ID AAV61450 standard; DNA: 3546 BP.  
 AC AAV61450;  
 XX  
 DT 02-FEB-1999 (first entry)  
 DE Human Clock gene.  
 KW Clock gene; circadian rhythm; human; jet lag; sleep-wake disorder;  
 KM seasonal affective disorder; cancer; diagnosis; therapy; ds.  
 XX  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT CDS 418..2958  
 FT /\*tag= a  
 PN WO9840514-A1.  
 PD 17-SEP-1998.  
 PF 13-MAR-1998; 98MO-US05114.  
 PR 30-JUN-1997; 97US-0885291.  
 PR 13-MAR-1997; 97US-0816693.  
 XX  
 PA (NOON ) UNIV NORTHWESTERN.  
 XX  
 PI Pinto LH, Takahashi JS, Turek F;  
 DR MPI: 1998-520828/44.  
 DR P-PSDB; AAW79533.  
 XX  
 PT Novel mouse and human circadian rhythm gene, clock - useful for  
 PT treating e.g. jet-lag, sleep-wake disorders, abnormal cell division,  
 PT etc  
 PS Claim 5; Fig 14.1-14.3; 154pp; English.  
 CC This is the nucleotide sequence of the human Clock gene that codes  
 CC for a putative 846-amino acid polypeptide (see AAW79533) involved in  
 CC the regulation of circadian rhythm. To isolate the gene, several  
 CC human clones identified in the NCBI database by end-sequence  
 CC analysis were sequenced, and a human hypothalamus cDNA library was  
 CC screened to isolate novel clones that hybridised with a probe of  
 CC the mouse Clock gene (see AAV61401). Further DNA sequence alignments  
 CC to the transcript of the mouse Clock gene revealed that a consensus  
 CC sequence from the aggregate of EST and hypothalamic clones extended  
 CC through the gene's entire coding region and into much of its  
 CC flanking 5' and 3' untranslated regions. The Clock gene regulates  
 CC at least 2 fundamental properties of the circadian clock system:  
 CC the intrinsic circadian period and the persistence of circadian  
 CC rhythmicity. The invention provides isolated and purified CLOCK  
 CC polypeptides, polynucleotides (including antisense), vectors and  
 CC host cells. These can be used to treat disorders of altered or  
 CC disrupted circadian rhythms e.g. jet-lag, seasonal affective  
 CC disorder, sleep-wake cycle disorders such as mood state, stress,  
 CC neurological disorders, to regulate diet and food intake especially  
 CC for diabetes, to treat cardiovascular, respiratory, liver or

CC endocrine disorders, and for diagnosis and treatment of abnormal  
 CC cell division such as cancer.  
 XX  
 SQ Sequence 3546 BP; 1118 A; 772 C; 728 G; 928 T; 0 other;  
 Query Match 0.9%; Score 55; DB 19; Length 3546;  
 Best Local Similarity 55.5%; Pred. No. 0.00058;  
 Matches 106; Conservative 0; Mismatches 85; Indels 0; Gaps 0;  
 QY 496 AAGAGTGCCTGGGCGCTATGATGCTTGAAGCCCTTGATGGCTTCTTTAGTGAAC 555  
 Db 736 aatgaagagttacacaataatgtagagcctctgagtgcttttttaagaacatcag 795  
 QY 556 CTGGAAGCAACGTTGTGTTGTGCAGAGATGTGACACAGATATTAAGGTATTAACAA 615  
 Db 796 acagatggaagcataaataatgctcgcgagagcgttaactcattcaactgaacattacca 855  
 QY 616 GAAGAGCTGATGAACAAGATGATATAGCATCTTGGCATTTGGGACACACAGGAATTT 675  
 Db 856 tcgatctctgtgatacaaatatattatccaccagaagggaacattcaagagtc 915  
 QY 676 GTCAAAAACCT 686  
 Db 916 tataaataact 926  
 RESULT 33  
 AAX03456  
 ID AAX03456 standard; cDNA: 5715 BP.  
 AC AAX03456;  
 XX  
 DT 10-MAY-1999 (first entry)  
 DE Human HSCLOCK cDNA.  
 KW HSCLOCK; clock gene; circadian rhythm; body clock; human;  
 KM sleep disorder; jet lag; diagnosis; therapy; ss.  
 XX  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT CDS 252..2792  
 FT /\*tag= a  
 PN EP890578-A2.  
 PD 13-JAN-1999.  
 PF 26-JUN-1998; 98BP-0305065.  
 PR 08-JUL-1997; 97EP-0304996.  
 XX  
 PA (SMIR ) SMITHKLINE BEECHAM PLC.  
 PI Duckworth DM, Michalovich D;  
 DR MPI: 1999-072834/07.  
 DR P-PSDB; AAW84565.  
 XX  
 PT New uses of clock gene (HSCLOCK) polypeptides and polynucleotides -  
 PT useful as diagnostic reagents and for treatment of sleep disorders,  
 PT jet lag and old age pathologies  
 PS Disclosure; Page 11-13; 22pp; English.  
 CC This polynucleotide comprises human HSCLOCK cDNA. The cDNA can be  
 CC isolated from a human brain cDNA library by expressed sequence tag  
 CC analysis, and can be used for producing HSCLOCK polypeptide (see  
 CC AAW83465) by recombinant methods. Sequence homology to the mouse  
 CC clock gene is about 87%. Recombinant host cells comprising an  
 CC expression system capable of producing a clock gene (HSCLOCK)



CC polypeptide are claimed. Also claimed are a method for treating a  
CC subject in need of enhanced activity or expression of HSCLOCK  
CC polypeptide by administering a HSCLOCK agonist or a polynucleotide  
CC encoding a HSCLOCK polypeptide; a method for treating a subject in  
CC need of having HSCLOCK activity or expression reduced by  
CC administering an antagonist of HSCLOCK, a nucleic acid that inhibits  
CC expression of a polynucleotide encoding HSCLOCK, or a polypeptide  
CC that competes with HSCLOCK polypeptide for its ligand, substrate or  
CC receptor; a process for diagnosing a disease, or susceptibility to  
CC disease, related to HSCLOCK expression of activity by determining  
CC the presence or absence of a mutation in the HSCLOCK gene, or  
CC determining the presence of level of HSCLOCK polypeptide expression;  
CC and a method for identifying compounds which inhibit (antagonise) or  
CC agonise the HSCLOCK polypeptide. HSCLOCK polypeptides and  
CC polynucleotides are useful for diagnosing and treating sleep  
CC disorders, jet lag, and pathologies that occur in advanced age  
CC related to an excess or lack of HSCLOCK polypeptide.  
XX  
50 Sequence 5715 BP; 1808 A; 1129 C; 1094 G; 1684 T; 0 other;

SQ Sequence 5715 BP; 1808 A; 1129 C; 1094 G; 1684 T; 0 other;

Query Match	0.98;	Score 55;	DB 20;	length 5715;
Best Local Similarity	55.58;	Pred. No.	0.0008;	
Matches 106;	Conservative	0;	Mismatches 85;	Indels 0;
				Gaps 0;

OY	496	AAGATGGGCGTGGGGCCGTATCATCTTGAGGCCCTTCATGGGTTCTCTTTGATGCAAC	555
Db	570	aatgaagagttcacacaataagtgtagaagcttcgtatggttcttttttagcaatcatg	629
OY	556	CTGGAAGGCACAAGTTGTGTTTGTGCAGAGATGTGACAGATATTAAAGGTATAACCAA	615
Db	630	acagatggaaacatatatatatgtctcgcgaagtgtaacttcattcacttgtaaacattacca	689
OY	616	GAAAGACTGATGAACAAAAGTATTAATPAGCATCTTGCAATGTTGGGGACCACACGGAATTT	675
Db	690	tctgatcttgygatlcaaagatatattaatttatccagaaggysnaacatcagaagtt	749
OY	676	GTCAAAAACCT	686
Db	750	tataaatatct	760

```

RESULT 34
ABL32623
ID ABL32623 standard; DNA; 6161 BP

```

AC ABL32623;

DT 26-MAR-2002 (first entry)

Human immune system associated gene SEQ ID NO: 596.

KW Human, immune system disease; cytosine methylation; antihistaminic;  
KW antiarteriosclerotic; antianemic; cyostatic; neotropic;  
KW neuroprotective; anti-HIV; anticoagulant; ophthalmological;  
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;  
KW antinflammatory; cancer; eye disease; arteriosclerosis; anaemia;  
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;  
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;  
KW gene; ds.

OS Homo sapiens.  
XX  
PN W0200200928-A2.  
XX  
PD 03-JAN-2002.  
XX  
XX 02-JUL-2001; 2001WO-EP07537  
PF  
XX 30-JUN-2000; 2000DE-1032529  
PR 01-SEP-2000; 2000DE-1043826  
XX  
XX (EPIC-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K;  
PI  
XX WPI; 2002-130909/17.  
DR

PT	Nucleic acid comprising fragment of chemically modified gene, useful
PT	for diagnosis and treatment of diseases associated with abnormal
PT	cytosine methylation -
XX	
PS	Claim 1; SEQ ID NO 596; 32pp + Sequence Listing; German.

The present invention provides a number of human immune system associated genes which are modified by the methylation of cytosines. The sequences can be used in the diagnosis and treatment of immune system disorders, including eye diseases such as retinopathy, neovascular glaucoma and macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis, rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel diseases. The present sequence is a gene of the invention.

50 Sequence 6161 BP; 1929 A; 50 C; 1116 G; 3065 T; 1 other;

Query Match	0.9%;	Score 55;	DB 24;	Length 6161;
Best Local Similarity	47.8%;	Pred. No. 0.00084;		
Matches 189; Conservative	0;	Mismatches 205;	Indels 1;	Gaps 1;

QY	5723	TTTTTATTTTGTGTTTTTACAGGCGACTAGAACCGAGTTTGTGAGGTACATAAAT	5782
Db	3738	ttttttgttgatcgatgaagaagtttttatagggatctaaatlttttagagaagaagatltt	3797
QY	5783	CCAAATTATATGTAAACAGCAATAATTTTAAGTGTGAGAACTTATGTGT -TAAATTGAT	5841
Db	3798	tttttttttataaggtatgtaaaatltttagaatgtaatttaagtggtttgtatltttat	3855
QY	5842	AATTTTGTGAGGATACATATGTGGAATTGACTCAAAAATGAGGTACTTAGATTTAA	5901
Db	3858	atttgctgttagcttggaaatctgtaattlttttaagaataaataaataatgattttaa	3917
QY	5902	ATTAGATTTCTCATPACCAATGTCCTTAAAGTGTTTGTAAAGATTCATATGCCTTG	5961
Db	3918	aaaaataatcttggaaataaataaataatcogtatatttaataaagtaataatttat	3977
QY	5962	ATTAGACCTAAATTTGTGAGACTTAAGACTTTTATTTTCAACCTGTGATTCGCTTAT	6021
Db	3978	tgatttagaagtttttaagttataataatagtaattgtttattattattattattataag	4037
QY	6022	AAGTCATTTATCTATCTATATATGATATGACGCCGCTGTAGAACCAATTCTGATTTTA	6081
Db	4038	tattataagtagatattataatttagagaatcttttggaaatgigtttttttttttt	4097
QY	6082	TATGTATTTATCTTTCTTATGACCTTAGAAG	6116
Db	4098	tagtaataaagtaattttaaattgatatataaag	4132

RESULT	35
AAx58984	
ID	AAx58984 standard; cDNA; 1581 BP.

AC AAX58984

DT 23-AUG-1999 (first entry)

Human transcription regulator MOP5 partial cDNA.

**KW** MOP5; member of the PAS superfamily; bHLH-PAS; human; transcription regulator; circadian signal transduction; ss.

	OS	XX	FH	FT	Key, CDS	Location/Qualifiers
						2.1447

```

FT      /*tag= a
FT      /partial
PN      WO928464-A2.
XX      10-JUN-1999.
XX      27-NOV-1998; 98WO-US25314.
XX      28-NOV-1997; 97US-0066863.
XX      (WISC ) WISCONSIN ALUMNI RES FOUND.
PA      Bradfield CA, Gu YZ, Hogenesch JB;
PI      WPI: 1999-371120/31.
DR      P-PSDB; AAY06293.
XX      Developmental signal transduction associated proteins
PT      Claim 6; Page 97-98; 106pp; English.
XX      This is the nucleotide sequence of a partial cDNA encoding MOP5 (see
CC      AAY06293), a novel member of the PAS superfamily, where PAS stands
CC      for PER/ARNT/SIM domains. The cDNA was identified in an iterative
CC      search of human ESTs designed to identify basic-helix-loop-helix-PAS
CC      (bHLH-PAS) proteins that interact with either the Ah receptor (AHR)
CC      or the Ah receptor nuclear translocator (ARNT). To obtain extended
CC      open reading frames for each EST, an anchored-PCR strategy was used
CC      to amplify additional flanking sequences from a commercial HepG2
CC      cDNA library. The invention provides novel MOP5 2'-9' nucleic acids
CC      (see AAS5881-88) and proteins (see AAY06289-97). These are useful in a
CC      variety of research, diagnostic and therapeutic applications.
CC      Several of the MOP5 are alpha-class hypoxia-inducible factors.
CC      Others are involved in circadian signal transduction.
XX      Sequence 1581 BP; 276 A; 560 C; 494 G; 251 T; 0 other;
SQ
Query Match      0.9%; Score 54.2; DB 20; Length 1581;
Best Local Similarity 55.6%; Pred. No. 0.00055;
Matches 104; Conservative 0; Mismatches 83; Indels 0; Gaps 0;
QY      505 CTGGGGCCATGATGCTTGAGGCCCTTGATGGTCTCTTTGTAGTGAACCTGGAAGC 564
DB      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      86 ctggggggtcacatcttcgagcagctccgtagtggttgcgtgccttgaaacagga 145
QY      565 AACGTGTGTTTGTGACGAGATGTGACACATATCTAAGTATACCAAGAAGCTG 624
DB      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      146 aaatctctacatctcagagacagctccatctatcttggtctctcagagtgagatg 205
QY      625 ATGAAACAAAAGTGTATATAGCATCTTGACATGTTGGGACACACAGCAATTGTCAAAAAC 684
DB      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      206 acggcgagcagcgctctgcactacatcacccctgggaccactcagagtgctggagca 265
QY      685 CTGCTGC 691
DB      ||| | |
DB      266 ctggggc 272

```

## RESULT 36

AAFI8177 standard; DNA; 4260 BP.

AAFI8177;

14-MAR-2001 (first entry)

Lung cancer associated polynucleotide sequence SEQ ID 196.

Human; lung cancer associated protein; neuroprotective; cytostatic;

cardioactive; immunomodulatory; muscular active; vulnerrary; gastrointestinal; nephrotoxic; antiinfective; gynecological;

```

KW      antibacterial; diagnosis; neural disorder; immune disorder; reproductive;
KW      proliferative disorder; wound healing; infectious disease; ds.
XX      Homo sapiens.
OS      WO200055180-A2.
XX      21-SEP-2000.
XX      08-MAR-2000; 2000WO-US05918.
XX      12-MAR-1999; 99US-0124270.
XX      (HUMA-) HUMAN GENOME SCI INC.
PA      (ROSE/) ROSEN C A.
XX      Ruben SM.
XX      WPI: 2000-587514/55.
DR      P-PSDB; AAB58301.
XX      Lung cancer associated gene sequences, referred to as lung cancer
PT      antigens, useful for treatment, prevention, and diagnosis of disorders
PT      such as lung cancer -
XX      Claim 1; Page 661-663; 1425pp; English.
XX      Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer
CC      associated proteins represented in AAB58106 - AAB55548. Lung cancer
CC      associated proteins and polynucleotide sequences, their agonists, and
CC      antagonists may have neuroprotective; cytostatic; cardioactive;
CC      immunomodulatory; muscular active general; vulnerrary; gastrointestinal
CC      general; nephrotoxic; antiinfective; gynecological; or antibacterial
CC      activity. The invention also includes antibodies specific for the
CC      protein or polynucleotide sequences. The lung cancer associated
CC      polynucleotide sequences may be used for detection of lung cancer,
CC      chromosome identification, as chromosome markers, and for numerous other
CC      diagnostic or research purposes. The proteins may be used to treat
CC      disorders such as neural, immune, muscular, reproductive,
CC      gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
CC      disorders. The proteins may also be used in the treatment of wounds and
CC      infectious diseases. Polynucleotide sequences AAF18425 - AAF18433 and
CC      peptide AAB58549 are used in the course of the invention for the
CC      identification and characterisation of the polynucleotide and protein
CC      sequences.
SQ      Sequence 4260 BP; 1297 A; 993 C; 887 G; 1077 T; 6 other;
Query Match      0.9%; Score 53.8; DB 21; Length 4260;
Best Local Similarity 52.0%; Pred. No. 0.0014;
Matches 250; Conservative 0; Mismatches 212; Indels 19; Gaps 5;
QY      4068 TCCAACTACGGAATAGTACGACCTGATCCAGGCTTTACTGGGCTACGACTGCCCA 4127
DB      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      1529 tccagagcaggaatggttccccaagtggagccaacttgcctcactctaaagccttg 1588
QY      4128 GAGCCCACTTATGTCACCCCGAATGGCAGCATACAGAGCCCATGATGCAACAGTCTCA 4187
DB      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      1589 gagctcatggtgcgagtccaatccctctcctcagagttctctcctcagcaaacac 1648
QY      4188 GGGCAACCCAGCCTATGAGCCGCCCTCCGACATAAATGATGGGCGAGGGAATGGG 4247
DB      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      1649 acctgcctccgggtatcagtcaccca--gacatgaagccttgacgaagagcgata-- 1703
QY      4248 CGAAGACAGCATGTTTCCAGCAGATCCACACACATTGGGACAGCAACACACAG 4307
DB      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      1704 -ggaaacaacaatggttccagctgtcagacacagccacgcctgcacagccagg 1762
QY      4308 CATGTACAGTAAACATGAAATCATGATGTCATGCGGACCAACACAGGTGCGATGAG 4367
DB      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      1763 agtatcc--aacacatgagcatcacggttccatcagg-----cagtggaataac 1810

```

OY		4368	CAGCATGAACCAAGATGACACAGGACAGATCAGTCACTAAGTACTCCGTGTCATTACGC	4427
OY				
Db		1811	gaatttcagaacaatgaaacccaatgatggccccagatgacagttagctctttgcagatggcc	1870
OY		4428	AAGCGTGTCCTCCATGAGGGGCC - GACCAGAGTTTAATCATCTCGTCGAGGGGAGGCAACC	4486
Db		1871	aggatgaactcgtgtgcccggaagcaataaagatgacctccgcacatggaacacacagccc	1930
OY		4487	TGTTCCCAACACGCTGCTCGAATGATGATTAAAGCAGGAGGAGACACAACAGCA	4546
Db		1931	tctactgcaaccagctctcatcacctgcacttctcaaaaagagcaagatggaaccag	1990
OY		4547	A 4547	
Db		1991	a 1991	
RESULT	37			
ID	AS45298		standard; DNA; 6665 BP.	
XX	AS45298;			
DT	18-DEC-2001	(first entry)		
DE		Chemically pretreated genomic DNA associated with cell cycle #2.		
XX				
KW		Cell cycle; human; CPG dinucleotide; cytosine methylation; HIV; aging;		
KW		human immunodeficiency virus; neurodegenerative disorder; solid tumour;		
KW		graft-versus-host disease; glomerular disease; Lewy body disease; cancer;		
KW		arthritis; arteriosclerosis; anti-HIV; neuroprotective; antiarthritic;		
KV		immunosuppressive; antitumour; cytostatic; antiarteriosclerotic; ds;		
OS	Homo sapiens.			
PN	WO200168911-A2.			
PD	20-SEP-2001.			
PE	15-MAR-2001; 2001WO-EP02945.			
PR	15-MAR-2000; 2000DE-1013847.			
PR	06-APR-2000; 2000DE-1019058.			
PR	07-APR-2000; 2000DE-1019173.			
PR	30-JUN-2000; 2000DE-1032529.			
PR	01-SEP-2000; 2000DE-1043826.			
PA	(EPID-) EPIGENOMICS AG.			
PI	Olek A, Plegenbrock C, Berlin K;			
DR	WPI: 2001-602751/68.			
PT		Designing primers and probes for analysing diseases associated with		
PT		cytosine methylation state e.g. arthritis, cancer, aging,		
PT		arteriosclerosis comprising fragments of chemically modified genes		
PT		associated with cell cycle -		
PS	Claim 1; SEQ ID No 3; 28pp; English.			
CC		Sequences AS45296-AS45520 represent chemically pretreated genomic DNA		
CC		molecules associated with the cell cycle and specific PCR primers of the		
CC		invention. The sequences are useful for detecting the methylation state		
CC		of all CpG dinucleotides in a sequence and therefore for analysing		
CC		associated diseases. By analysing cytosine methylations in the pretreated		
CC		DNM, genetic and/or epigenetic parameters for the diagnosis and therapy		
CC		of existing diseases or the predisposition to specific diseases can be		
CC		ascertained. The parameters may be compared to another set of genetic		
CC		and/or epigenetic parameters, the differences serving as basis for		
CC		diagnosis and/or prognosis events which are disadvantageous to patients.		
CC		The sequences of the invention are useful for the diagnosis and therapy		
CC		of HIV infection, neurodegenerative disorders, graft-versus-host disease,		

[illegible]



KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;  
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;  
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;  
KW gene; ds.

OS Homo sapiens.

XX MO200200928-A2.

XX 03-JAN-2002.

XX 02-JUL-2001; 2001WO-EP07537.

XX 30-JUN-2000; 2000DE-1032529.

XX 01-SEP-2000; 2000DE-1043826.

XX (EPIC-) EPIDENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K;

XX WPI; 2002-130909/17.

XX Nucleic acid comprising fragment of chemically modified gene, useful  
PT for diagnosis and treatment of diseases associated with abnormal  
PT cytosine methylation

XX cytosine methylation

PS Claim 1; SEQ ID NO 1280; 32pp + Sequence listing; German.

CC The present invention provides a number of human immune system associated  
CC genes which are modified by the methylation of cytosines. The sequences  
CC can be used in the diagnosis and treatment of immune system disorders,  
CC including eye diseases such as retinopathy, neovascular glaucoma and  
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid  
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel  
CC diseases. The present sequence is a gene of the invention.

XX Sequence 6175 BP; 1506 A; 154 C; 1357 G; 3158 T; 0 other;

Query Match 0.9%; Score 53.4; DB 24; Length 6175;

Best Local Similarity 48.5%; Pred. No. 0.0024;

Matches 205; Conservative 0; Mismatches 216; Indels 2; Gaps 2;

OY 5672 TGTGTTAAATGATAGACATATCAGACAGATTAACTGCGCATGTTTATT 5731

DB 4323 tgltaataattataaagtaattatttttataatagttataatagttata 4382

OY 5732 TTGTTTTTACAGCTGCTGACTAAGCCGAAAGTTTGTAGGTACATAAATCAATTAT 5791

DB 4383 ttatttataacglttataataacgtaatttttataattattattattatglttaa 4442

OY 5792 ATGTAACAAAGCAATTAATTAAGTGAGAACTATGTTTAAATGTAATTTTGT 5850

DB 4443 atataaagtaattattattattattattattataacgltataacgtaacgtattt 4502

OY 5851 GA-GGTATACATATTGTGGAATTCACCAAAATAGAGTACTCAGTATTAATTAAGTA 5909

DB 4503 gataataataatglttataataataataacgtaacglttataatgattataat 4562

OY 5910 TCTTCATAGCAATGCTCCCTAAAGTGTTTGTAAAGATATCAATGCTGATTAGAC 5969

DB 4563 tataatgltataataatgltgtaagtaacgtatgattgattataattat 4622

OY 5970 TAATTGTAGACTTAAGACTTTTATTTCTAAACCTTGATGCTGCTTAAAGTCATT 6029

DB 4623 ttatttgataattattattatttttataatgltatttttataatattat 4682

OY 6030 TATCTAATCTATATATATGACCGCGGTAGAACCAATCTGATTTTATATGTTTA 6089

DB 4683 ataagtaataagttataattataataatgattataattattataaataatglttata 4742

OY 6090 TAT 6092

DB 4743 tat 4745

Search completed: September 12, 2002, 09:56:35  
Job time: 21609 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 12, 2002, 03:38:46 ; Search time 10727.9 Seconds  
(without alignments)  
12008.337 Million cell updates/sec

Title: US-09-842-256-1  
Perfect score: 6156  
Sequence: 1 GCGCGCCGACGCTCGCTA.....CCACTTTATGTTGTTT 6156

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:\*

- 1: gb\_ha:\*
- 2: gb\_hgt:\*
- 3: gb\_in:\*
- 4: gb\_om:\*
- 5: gb\_ov:\*
- 6: gb\_pat:\*
- 7: gb\_ph:\*
- 8: gb\_pl:\*
- 9: gb\_pr:\*
- 10: gb\_ro:\*
- 11: gb\_sts:\*
- 12: gb\_sy:\*
- 13: gb\_un:\*
- 14: gb\_vl:\*
- 15: em\_ba:\*
- 16: em\_fun:\*
- 17: em\_hum:\*
- 18: em\_in:\*
- 19: em\_mu:\*
- 20: em\_om:\*
- 21: em\_or:\*
- 22: em\_ov:\*
- 23: em\_pat:\*
- 24: em\_ph:\*
- 25: em\_pl:\*
- 26: em\_ro:\*
- 27: em\_sts:\*
- 28: em\_un:\*
- 29: em\_vl:\*
- 30: em\_hgt\_hum:\*
- 31: em\_hgt\_inv:\*
- 32: em\_hgt\_other:\*
- 33: em\_hgtgo\_inv:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Query Match length	DB ID	Description
-----			

1	6154.4	100.0	6156	9	HSTIF2GEN	X97674 H.sapiens m
2	3846	62.5	4878	10	MMU39060	U39060 Mus musculu
3	3699.6	60.1	4771	10	AF000582	AF000582 Mus muscu
4	3593.2	58.4	4398	10	AF136943	AF136943 Rattus no
5	1961.4	31.9	4871	5	X1A243119	A1243119 Xenopus l
6	1592.8	25.9	86554	2	AC090280	AC090280 Homo sapi
7	1592.8	25.9	181403	2	AC084251	AC084251 Homo sapi
8	1591.2	25.8	71701	2	AF254984	AF254984 Homo sapi
9	1274.2	20.7	86554	2	AC090280	AC090280 Homo sapi
10	1274.2	20.7	129924	2	AF207954	AF207954 Homo sapi
11	1274.2	20.7	124630	9	AC090731	AC090731 Homo sapi
12	1238.4	20.1	4958	5	AF323989	AF323989 Danio rer
13	1071.2	17.4	110000	2	AC091288-0	AC091288 Mus muscu
14	1036	16.8	155871	2	AC095150	AC095150 Rattus no
15	564.2	9.2	4609	10	AF000581	AF000581 Mus muscu
16	553.4	9.0	4495	9	AF010227	AF010227 Homo sapi
17	542.8	8.8	6835	9	AF012108	AF012108 Homo sapi
18	533.6	8.7	4668	9	AF016031	AF016031 Homo sapi
19	524.6	8.5	110000	2	AC091288-0	AC091288 Mus muscu
20	524.6	8.5	223169	2	AC091248	AC091248 Mus muscu
21	517.6	8.3	4546	5	AF044080	AF044080 Xenopus l
22	510.6	8.4	6754	9	AF036892	AF036892 Homo sapi
23	454.8	7.4	110000	2	AC091288-1	Continuation (2 of
24	369.8	6.0	71701	2	AF254984	AF254984 Homo sapi
25	369.6	6.0	5169	10	MMU64828	U64828 Mus musculu
26	368	6.0	4752	10	MMU64606	U64606 Mus musculu
27	366.4	6.0	4752	10	MMU64606	U64606 Mus musculu
28	366.2	5.9	4664	9	HSJ000881	AJ000881 Homo sapi
29	366.2	5.9	4721	9	HSJ000882	AJ000882 Homo sapi
30	364.6	5.9	4660	9	HSU59302	U59302 Human stero
31	363.2	5.9	4547	9	HSU09061	U9061 Human stero
32	362.8	5.9	4527	5	AF443296	AF443296 Coturnix
33	339.8	5.5	399	6	AX351515	AX351515 Sequence
34	250	4.1	250	11	G15398	G15398 human STS S
35	237.4	3.9	104593	2	AC091288-3	Continuation (4 of
36	195.4	3.2	4036	10	AF322224	AF322224 Rattus no
37	194	3.2	155871	2	AC095150	AC095150 Rattus no
38	192	3.1	70606	2	AC109202	AC109202 Mus muscu
39	154.4	2.5	223169	2	AC091248	AC091248 Mus muscu
40	150.6	2.4	1367	2	AK021838	AK021838 Homo sapi
41	136.6	2.2	110000	2	AC091288-1	Continuation (2 of
42	132.2	2.1	207810	2	AC091158	AC091158 Mus muscu
43	102.8	1.7	86453	9	HS1049316	AL034418 Human DNA
44	101.4	1.6	188741	2	AC013459	AC013459 Homo sapi
45	95.4	1.5	170936	10	AL589873	AL589873 Mouse DNA

#### ALIGNMENTS

RESULT 1  
HSTIF2GEN LOCUS 6156 bp mRNA linear PRI 10-MAR-1997  
DEFINITION H.sapiens mRNA for transcriptional intermediary factor 2.  
X97674.1  
VERSION X97674.1 GI:1877214  
KEYWORDS alternatively spliced; nuclear receptor coactivator; TIF2 gene; transcriptional mediator.  
SOURCE human.

#### ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

#### REFERENCE

Voegel,J.-J., Heine,M.J., Zechel,C., Chambon,P. and Gronemeyer,H.  
TIF2, a 160 kDa transcriptional mediator for the ligand-dependent  
activation function AF-2 of nuclear receptors  
EMBO J. 15 (14), 3667-3675 (1996)

#### JOURNAL

EMBO J. 15 (14), 3667-3675 (1996)

#### MEDLINE

96312964

#### REFERENCE

2 (bases 1 to 6156)

#### AUTHORS

Voegel,J.-J.

#### TITLE

Submitted (22-APR-1996) J.J. Voegel, IGBMC Inst.de Genet.et

#### JOURNAL

Biol.Mol.et Cell., CNRS-INSERM-Univ.Louis Pasteur, B.P.163, C.U. de

Strasbourg, F-67404 ILLKIRCH CEDEX, FRANCE  
Revised by author 25-JUL-96 and 10-MAR-97  
On Mar 11, 1997 this sequence version replaced gi:1490315.  
COMMENT  
Related sequences U33060, U40396.

FEATURES  
Location/Qualifiers

Source

1. 6156

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_lib="lambdaexlox Ref.No.56"

163..4557

/gene="TF2"

163..4557

/gene="TF2"

/function="transcriptional mediator for ligand-dependent  
activation function AF-2 of nuclear receptors"

/codon\_start=1

/product="transcriptional intermediary factor 2"

/protein\_id="CA66263.1"

/db\_xref="GI:1877215"

/translation="MSGGENTSDPSRAETRRKRECPDQIGSPKRNTERKREDEENK  
YIEELAELEAFENNDIDNENFDPKCAIKETRYKROIKEDEKAAANIDYOKSDY  
SSTGGVLDKALGPMLEALDGFPPVNLBGNVYVSENTOYLYRNOBELMNSVY  
SLHVDHTEPYKNLPSKIVNGSGSGEPKRNSTTMCRLVYKLPDYEEGHNDQ  
EAHQKETEYQCAVSPKSIKEGEDJQCLICVARVPMKEPVLPSSEFTTRDL  
OGKITSLDSTYKAAKPGMEDLVKRCIQKFAHGESEVYAKRHHEVYLRGLAFS  
QIYRFLSDGTVAQTKSLRSQTNTPOLVLSHMLAREQNVNMPDLTGTMG  
KLPNLSNSPAHQAICSGNPEODMTLSINPFIEMHREQNGMPGRGSGGMHV  
SGMOATPGGSNYALKMNSPSSSGMNGMPTSMSPHRMSPVAGSPAPRPSQS  
PAGSLHSPYGCSTGNSSTYSSINLALQALSEGYSGLSLAPDLKMGILNSP  
VYNNPPPLSKMSLDKDCFGLEPSEGTGOAESCHPGEKDETNDNLPVAVSE  
RDEKOSRLHDSKQKRLQLQTLTKSDQMEPSLASSLSDTNDSTGLPSSTHGT  
LAEKHRLHRLDQSSPYDLAKLTAEATGKDLSSSSSTAEGSEVYIKOEVPSPKK  
EVALREYLDKDDTDIGLEITPKLERLSDKDPASNTKILAMKTEKEMSPEDQ  
PSELDNLEETLIDLONSLPOLFPDTRGAPGASVDKQAIINDLMOLTAENSPYVY  
GAOKTALRTISOSTFNNRPGOLRLPNLNLDTLQSPTEGAPPPYLRNSPYVI  
POPGMNGMGIGNOCNLNSTGMSASRTMPSGEMAPQSSAVRYTCAATTSAM  
NRPVOGMI RNPAASIPMRPSSQPGQRTLSQVMNIGSELEMMNGPQY500QAP  
NOTAPWPEBILIDQAFASQONRPGSSPDDILCPHPAESPSDEGALIDQYLALR  
NFDGLEIDRALGIPELVSOQAVDEPQSSODSNLMLOKAPVFQYQASQOMAG  
SYSPMODPNEHMGORSPYATLRMOPRGLRPLGVALGVALONHLOHRLAOQOR  
PLMNQISVSNVNLTRPGVPTQAPINAMLOARETINOHRLROQMDOQOORT  
LMMRGGLMTPSKVAPSGMPRTMSNPIRQAAQDFPPPTNGISQDPDQGTGAT  
POSPLMSPRMAHTQSPMQSOANPAYQAPSDINGAAGNMGNSNF500SPHFQOQ  
AATSMYSNNMANTVSMNTNGMSNMONTGOISMTSVTSVPTGSLSMGPEQVNDPA  
LRGNLFPNQLPGMDIMIKOEGDTRKYC"

2768..2974

/gene="TF2"

/note="putative alternatively spliced region"

BASE COUNT 1799 a 1493 c 1406 g 1458 t

ORIGIN

misc\_feature

2768..2974

/gene="TF2"

/note="putative alternatively spliced region"

BASE COUNT 1799 a 1493 c 1406 g 1458 t

ORIGIN

Query Match 100.0%: Score 6154.4: DB 9: Length 6156;

Best Local Similarity 100.0%: Pred.No.0:

Matches 6155: Conservative 0: Mismatches 1: Indels 0: Gaps 0:

1 GGGGGCGCAGCCTCGGCTACAGCTTCGGGGGGAAGTCAAGCCGCGACCGCGCA 60

1 GGGGGCGCAGCCTCGGCTACAGCTTCGGGGGGAAGTCAAGCCGCGACCGCGCA 60

1 GGGGGCGCAGCCTCGGCTACAGCTTCGGGGGGAAGTCAAGCCGCGACCGCGCA 60

1 GGGGGCGCAGCCTCGGCTACAGCTTCGGGGGGAAGTCAAGCCGCGACCGCGCA 60

1 GGGGGCGCAGCCTCGGCTACAGCTTCGGGGGGAAGTCAAGCCGCGACCGCGCA 60

1 GGGGGCGCAGCCTCGGCTACAGCTTCGGGGGGAAGTCAAGCCGCGACCGCGCA 60

1 GGGGGCGCAGCCTCGGCTACAGCTTCGGGGGGAAGTCAAGCCGCGACCGCGCA 60

1 GGGGGCGCAGCCTCGGCTACAGCTTCGGGGGGAAGTCAAGCCGCGACCGCGCA 60

1 GGGGGCGCAGCCTCGGCTACAGCTTCGGGGGGAAGTCAAGCCGCGACCGCGCA 60

1 GGGGGCGCAGCCTCGGCTACAGCTTCGGGGGGAAGTCAAGCCGCGACCGCGCA 60

1 GGGGGCGCAGCCTCGGCTACAGCTTCGGGGGGAAGTCAAGCCGCGACCGCGCA 60

1 GGGGGCGCAGCCTCGGCTACAGCTTCGGGGGGAAGTCAAGCCGCGACCGCGCA 60

1 GGGGGCGCAGCCTCGGCTACAGCTTCGGGGGGAAGTCAAGCCGCGACCGCGCA 60

1 GGGGGCGCAGCCTCGGCTACAGCTTCGGGGGGAAGTCAAGCCGCGACCGCGCA 60

1 GGGGGCGCAGCCTCGGCTACAGCTTCGGGGGGAAGTCAAGCCGCGACCGCGCA 60

1 GGGGGCGCAGCCTCGGCTACAGCTTCGGGGGGAAGTCAAGCCGCGACCGCGCA 60

1 GGGGGCGCAGCCTCGGCTACAGCTTCGGGGGGAAGTCAAGCCGCGACCGCGCA 60





Db	3481	AGCCAGAGCCCAAGCAGTAGATGCATCCAGAACAGTTCTCAAGTCAGAGATTCCAAATCATATGCTG	3540
Qy	3541	GAGCAGAGAGGCGCCGTTTTCCTCCACAGCAGTATGCATCTCAGGCAACAAATGGCCAGGCT	3600
Db	3541	GAGCAGAGAGGCGCCGTTTTCCTCCACAGCAGTATGCATCTCAGGCAACAAATGGCCAGGCT	3600
Qy	3601	AGCTATTCTCCCATGACAGATGCCAAACTTTCACACCATTGGGACAGCGGCTTAGTTATGCC	3660
Db	3601	AGCTATTCTCCCATGACAGATGCCAAACTTTCACACCATTGGGACAGCGGCTTAGTTATGCC	3660
Qy	3661	ACACTCCGTTATGCAGGCCACAGCGGGGCTCAGGGCCAGGGGCTGTGTGCAGACACAGCA	3720
Db	3661	ACACTCCGTTATGCAGGCCACAGCGGGGCTCAGGGCCAGGGGCTGTGTGCAGACACAGCA	3720
Qy	3721	AATCACTTAAGACTTCAACTTCAGATTCGCTCTCAAGCACACAGAAATCGCCAGCCACTT	3780
Db	3721	AATCACTTAAGACTTCAACTTCAGATTCGCTCTCAAGCACACAGAAATCGCCAGCCACTT	3780
Qy	3781	ATGGAATCAAAATCAGCAATGTTTCCAAATGTGAACCTTGACTCTGAGGCTTGAGTACCA	3840
Db	3781	ATGGAATCAAAATCAGCAATGTTTCCAAATGTGAACCTTGACTCTGAGGCTTGAGTACCA	3840
Qy	3841	CAGGACCTCTTATATGCAGAGATGTGGGCCAGAGACAGAGGGAATCTGAACACAGAT	3900
Db	3841	CAGGACCTCTTATATGCAGAGATGTGGGCCAGAGACAGAGGGAATCTGAACACAGAT	3900
Qy	3901	CTTCGACAGAGACAAATGCATTCAGCAACAGCAAGTTTCAGCAACGAACTTTGATGATGA	3960
Db	3901	CTTCGACAGAGACAAATGCATTCAGCAACAGCAAGTTTCAGCAACGAACTTTGATGATGA	3960
Qy	3961	GGACAAAGGTTGAATATGACACCAACAGATGTGGTCTCTAGTGTGATGCCAGCAACTATG	4020
Db	3961	GGACAAAGGTTGAATATGACACCAACAGATGTGGTCTCTAGTGTGATGCCAGCAACTATG	4020
Qy	4021	AGCAACCTCGGATTTCCCCAGGGCAATGCACAGAGATTCCATTTCTCTCAACACAGGA	4080
Db	4021	AGCAACCTCGGATTTCCCCAGGGCAATGCACAGAGATTCCATTTCTCTCAACACAGGA	4080
Qy	4081	ATAAGTCAGCAACCTGCATCCAGGCTTTACTGTGGGGCTACGACTCCCCAGAGGCCACTTATG	4140
Db	4081	ATAAGTCAGCAACCTGCATCCAGGCTTTACTGTGGGGCTACGACTCCCCAGAGGCCACTTATG	4140
Qy	4141	TCACCCCGAATGGCACAATACACAGAGTCCCATGTATGCAACAAGTCTCAGAGCCACCCAGCC	4200
Db	4141	TCACCCCGAATGGCACAATACACAGAGTCCCATGTATGCAACAAGTCTCAGAGCCACCCAGCC	4200
Qy	4201	TATCAGGCCCCCTCCGATCAATTAATGTGATGGGCGAGGGGAATATGGGGGGAACAGCATG	4260
Db	4201	TATCAGGCCCCCTCCGATCAATTAATGTGATGGGCGAGGGGAATATGGGGGGAACAGCATG	4260
Qy	4261	TTTTTCCAGAGATCCCCACACACTTTTGGGCGAGCAAGCAAAACCCAGCATGTACAGTAAAC	4320
Db	4261	TTTTTCCAGAGATCCCCACACACTTTTGGGCGAGCAAGCAAAACCCAGCATGTACAGTAAAC	4320
Qy	4321	AACATGAACATCAATGTGTCCATATGGGACCAACACAGTGGCATAGCAGATGAACAG	4380
Db	4321	AACATGAACATCAATGTGTCCATATGGGACCAACACAGTGGCATAGCAGATGAACAG	4380
Qy	4381	ATGACAGACAGATCACACATGACCTCAGAGACCTCCGTGTACGTGAGGGCTGGCTGCC	4440
Db	4381	ATGACAGACAGATCACACATGACCTCAGAGACCTCCGTGTACGTGAGGGCTGGCTGCC	4440
Qy	4441	ATGAGGTCCCGACAGAGTTATATGATCTCGCTGTAGGGGAGGCACACTTCTCCAAACAG	4500
Db	4441	ATGAGGTCCCGACAGAGTTATATGATCTCGCTGTAGGGGAGGCACACTTCTCCAAACAG	4500
Qy	4501	CTGCGCTGAATGATATGATTAAGACAGAGGAGAGACAAACGGAATATTTGCTGACAC	4560
Db	4501	CTGCGCTGAATGATATGATTAAGACAGAGGAGAGACAAACGGAATATTTGCTGACAC	4560
Qy	4561	TGCTGAAGCCAGTTGCTTTCTTACGCTGACCGGGCTCCTGCTCAAAAACCTTCCAGCT	4620
Db	4561	TGCTGAAGCCAGTTGCTTTCTTACGCTGACCGGGCTCCTGCTCAAAAACCTTCCAGCT	4620

QY	4621	GGAGAGCTGTGTCTATTTTGTGTTTCAACCCAACTGACTCCGACCGGGTTCTGCTAGACAG	4680
Db	4621	GGAGAGCTGTGTCTATTTTGTGTTTCAACCCAACTGACTCCGACCGGGTTCTGCTAGACAG	4680
QY	4681	ACAGAGCCGGCCCGGTGTGCCAGGGTGGCGCTCCAGTCGGCTGGGAGAGAGAGCTGGCT	4748
Db	4681	ACAGAGCCGGCCCGGTGTGCCAGGGTGGCGCTCCAGTCGGCTGGGAGAGAGAGCTGGCT	4740
QY	4741	CTTCTCTTGACAGTCTGAAGCTCGCATCCAGACAGTCGGCTCAGTCTGTTCCCTGCATTC	4800
Db	4741	CTTCTCTTGACAGTCTGAAGCTCGCATCCAGACAGTCGGCTCAGTCTGTTCCCTGCATTC	4800
QY	4801	CCTTAGAGCAACTAGATATCTCTCTCCGCCAAGTAATGTGACAGGCCAATTTCAATACC	4860
Db	4801	CCTTAGAGCAACTAGATATCTCTCTCCGCCAAGTAATGTGACAGGCCAATTTCAATACC	4860
QY	4861	ATGTCAAGTGAATGATATTTTAAATGTATGTATTAAGAGAAACATAGCTCTGTCTCTT	4920
Db	4861	ATGTCAAGTGAATGATATTTTAAATGTATGTATTAAGAGAAACATAGCTCTGTCTCTT	4920
QY	4921	CCTGTGGGTTCCAGACAGTGGTTCTGTGTTTTCCTCGGCTAACAGTCTAGTGC	4980
Db	4921	CCTGTGGGTTCCAGACAGTGGTTCTGTGTTTTCCTCGGCTAACAGTCTAGTGC	4980
QY	4981	AAAGATTAAGATTTTATCTGGGGGAAAGAAAGAAATTTTAAAAATTAATTAATTAAGAT	5040
Db	4981	AAAGATTAAGATTTTATCTGGGGGAAAGAAAGAAATTTTAAAAATTAATTAATTAAGAT	5040
QY	5041	GTTTAAAGCTTAAGCCCTGGAATTTTGGATGGAGACAGACAGACCGGTGACAGCGCTGT	5100
Db	5041	GTTTAAAGCTTAAGCCCTGGAATTTTGGATGGAGACAGACAGACCGGTGACAGCGCTGT	5100
QY	5101	ATTATACAGACACACCAGCTGCTGAAGACCAACAAAGTACAGTGTATCTTAAGAAAG	5160
Db	5101	ATTATACAGACACACCAGCTGCTGAAGACCAACAAAGTACAGTGTATCTTAAGAAAG	5160
QY	5161	TCCTAAAGCATTGTTGGAAAAGTCTCCAGTTACTAACAAGATGAAGAAAGACCTGTGAG	5220
Db	5161	TCCTAAAGCATTGTTGGAAAAGTCTCCAGTTACTAACAAGATGAAGAAAGACCTGTGAG	5220
QY	5221	AGGGGTGTAAACATTACCAATATTTTCTCTGTTTTCTTGTGTTAAACCAAGACGG	5280
Db	5221	AGGGGTGTAAACATTACCAATATTTTCTCTGTTTTCTTGTGTTAAACCAAGACGG	5280
QY	5281	TTTACCTGTAATCATGAATTAAGAAAGAAATATTTTCATTTCAATTAATTAAGTCCCTTTAG	5340
Db	5281	TTTACCTGTAATCATGAATTAAGAAAGAAATATTTTCATTTCAATTAATTAAGTCCCTTTAG	5340
QY	5341	TTTATACAGACAGTTGAATACAGATCTCTCTCTCCACAGCCGATCTCTCCCTCC	5400
Db	5341	TTTATACAGACAGTTGAATACAGATCTCTCTCTCCACAGCCGATCTCTCCCTCC	5400
QY	5401	CCTCTCTCATCCCCATCACTCCCTATTTTCATTCCTTTTAAAAAATTAATTAAGCTAC	5460
Db	5401	CCTCTCTCATCCCCATCACTCCCTATTTTCATTCCTTTTAAAAAATTAATTAAGCTAC	5460
QY	5461	AGAAACACAGGTAAAGCCCTTATTTCTTAATGTTTGGCAGCACTTACCAATTTGCTAA	5520
Db	5461	AGAAACACAGGTAAAGCCCTTATTTCTTAATGTTTGGCAGCACTTACCAATTTGCTAA	5520
QY	5521	GTAATTAATTTCAAAAAAATTAATGATTTACTGGCAAGGAAAGACAAAGTTAAAGC	5580
Db	5521	GTAATTAATTTCAAAAAAATTAATGATTTACTGGCAAGGAAAGACAAAGTTAAAGC	5580
QY	5581	TTGTATACCAATCGAGCTAAAGATACCTGCTTTTGGAGACATGTTTATCTGTTCCCCAC	5640
Db	5581	TTGTATACCAATCGAGCTAAAGATACCTGCTTTTGGAGACATGTTTATCTGTTCCCCAC	5640
QY	5641	ACTCTGGCTCTCAAAATGGAGAAAGCCAGCTGTGTTTAAATTAATTAAGCATATACGA	5700
Db	5641	ACTCTGGCTCTCAAAATGGAGAAAGCCAGCTGTGTTTAAATTAATTAAGCATATACGA	5700



QY	301	GAGAACTTGCAGAGTATGATTTTGGCAAAATTTTAATATATAGCAACTTTAACTTAAA	360
Db	342	GAGGAGCTGGCCGAGTCTATCTTGGCAAACTTTAATGATATGCAAACTTCAACTTCAAA	4010
QY	361	CTGACAAATGTCAGTATTAAAAAGAAACTGTGAAGCAAAATTCGTACATCAAAAGACAA	4200
Db	402	CTGACAAATGTCGATCTTAAAAAAMCTGTGAAGCAGATCCGCCAGATCAAAAGACAA	4610
QY	421	GAGAAAGCAGCAGCTGCCAACATAGATGAGTGCAGAAAGTCAGATGTAATCCTTACAGG	4800
Db	462	GAGAAAGCAGCAGCTGCCAACATAGATGAGTGCAGAAAGTCAGATGTAATCCTTACAGG	5210
QY	481	CAGGGTGCATGAGCAAGAGATGCGGTGGGGCCCTGATGATGCTTGAAGGCCCTGATGGGTTTC	5420
Db	522	CAGGGTGCATGAGCAAGAGATGCGGTGGGGCCCTGATGATGCTTGAAGGCCCTGATGGGTTTC	5610
QY	541	TTCTTTGTAGTGAACCTGGAAGGCAACGCTGTGTTGTGTGCAGAAATGTCAGACAGTAT	6000
Db	582	TTCTTGTGTGTAACCTGGAAGGCAAGTGTGTGTTGTGTGCAGAAATGTCAGACAGTAT	6410
QY	601	CTAAGGTATTAACCAAGAAGAGCTGATGAAACAAAGTATATAGCATCTTGATGTTGGG	6600
Db	642	CTAAGGTATTAACCAAGAAGAGCTGATGAAACAAAGTGTCTACACATCTGTCAGATCTGGG	7010
QY	661	GACCAACAGGGAATTTGTCAAAAACCTGCTGCCAAAGTCTATAGTAATGGGGGACTCTTGG	7200
Db	702	GACCAACAGGGAATTTGTCAAAAACCTGCTGCCAAAGTCTATAGTAATGGGGGACTCTTGG	7610
QY	721	TCTGGCAACCTCCGAGCGGAGAACCCATACCTTCAATATGTGGATGTGTGTAACCT	7800
Db	762	TCTGGAGAACTCCCGAGCGGAGAACCCATACCTTCAACATGTGCCATGTGGTGAAGCCT	8210
QY	781	TTACCTGATTCGAAGAAGAGAGGTCATGATATACCAAGAGTCATCAGAAATATGAAACT	8400
Db	822	TTGCCAGATTCGAAGAAGAGAGGTCATGATATACCAAGAGTCATCAGAAATATGAAAGCG	8610
QY	841	ATGCAAGTCTTCGCTGCTCTCAACCAAAAGTCATCAAAAGAGAGAGAAAGATTTGTCAG	9000
Db	882	ATGCAAGTCTTCGCTGCTGCTCTCAAGCCCAAGTCATCAAAAGAGAGAGAAATTTGTCAG	9410
QY	901	TCTCTCTTGATTTGGCTGTGGCAAGAGATTTCCCATGAAAGAGAACCAATTTTCTCTCA	9600
Db	942	TCTCTCTTGATTTGGCTGTGGCAAGAGATTTCCCATGAAAGAGAACCAATTTTCTCTCA	10000
QY	961	TCAGAAATTTTACTACTGCGCAGAGATCTCCAAAGCAAGATACGTCCTGATATACAGC	10200
Db	1002	TCAGAAATCTTACCAACCCGCGCAGAGATCTCCAAAGCAAGATACGTCCTGATATACAGC	10600
QY	1021	ACCATGAGAGCAGCCATGTAACCAAGCTGGGAGAGACCTGTTAGAGAGGTATTTACAGAG	10800
Db	1062	ACCATGAGAGCAGCCATGTAACCAAGCTGGGAGAGATCTGTTAGAGAGATTCATTTACAGAG	11200
QY	1081	TTCCATGCGCAGATGAAGAGAGATGTGTGCTCATATCTTAAGAGGCAATCATGTAAGATA	11400
Db	1122	TTCCATGCGCAGATGAAGAGAGATGTGTGCTCATATCTTAAGAGGCAATCATGTAAGATA	11800
QY	1141	CTGAGACAGAGATGGCATTCATGCTCAAAATCTATGTTTCTTGTGTCGATGGCAGCTTT	12000
Db	1182	CTGAGACAGAGATGGCTTGGCTTACGTCAGATCTATGTTTCTTGTGTCGATGGCAGCTTC	12400
QY	1201	GTTTGTGCAACAAGCAAGAGAAAGTATACGTTCTTCAGAGCTACTATGAAACCTGCAACTT	12600
Db	1242	GTTTGTGCAACAAGCAAGAGAAAGTATACGTTCTTCAGAGCTACTATGAAACCTGCAACTT	13000
QY	1261	GTAATATCTTTACATATCTCTTACAGAGAGAGAGATGTGTGTGATGAATCCGAGATCTG	13200
Db	1302	GTAATATCTTTACATATCTCTTACAGAGAGAGAGATGTGTGTGATGAATCCGAGATCTG	13600
QY	1321	ACTGACAAAGATGGGAGAACCCATGATTCATTTAGTCTTAAACAGCCTGCCCATCAG	13800
Db	1362	ACTGACAAAGATGGGAGAACCCATGATTCATTTAGTCTTAAACAGCCTGCCCATCAG	14200

QY	1381	GCCTGTGCACTGGGAACCCAGGTGCAGAGCATACCTCAGTATGCAATATTAATTTTCCC	1440
Db	1422	GCCCTGTCACTGGGAACCCAGGTGCAGAGCATACCTCAGTATTAATTTTCCC	1481
QY	1441	ATTAATTTGGCCCAAGGAACAAATGGGATGCCCATGGGAGGTTTGGTGTCTGGGGGA	1500
Db	1482	ATGAATTTGGCCCAAGGAACAAATGGGATGCCCATGGGAGGTTTGGTGTCTGGGGGC	1541
QY	1501	ATGAACCATGTGTCAAGGATGCAGAACGACATCTCTAGGGTAGTAACTATGCACTCAAA	1560
Db	1542	ATGAACCATGTGTCAAGGATGCAGAACGACATCTCTAGGGTAGTAACTATGCACTCAAA	1601
QY	1561	ATGAACAGCCCCCTCAACAAGCAGCCCTGGCATATATCAGAGACGCCACATCTCCATT	1620
Db	1602	ATGAACAGCCCCCTCGCAAAAGAGCCCGGCGATGAACCCGGGGGCAACCCAGTCCCTGCTC	1661
QY	1621	TCACCAAGGCAATGCGATGAGCCCTGGAGTGGGTGGAGGCCCTCGAATCCACCACTGATG	1680
Db	1662	TCCCCAAGGCAAGCGATGAGCCCGGCGTGGTGGAGTCTCTGGCATCCCACTCCAGTATG	1721
QY	1681	TTTTTCCCTGCGAGAGAGCTTGTCATTTCCCTGTGTGGAGTTTGGACAGCAGCAAGAAATAGC	1740
Db	1722	TTTTTCCCTGCGAGAGAGCTTGTCATTTCCCTGTGTGGAGTTTGGACAGCAGCAGAAATATAGC	1781
QY	1741	CATAGTTATACCAACAGCTCCCTCATATGACATTCAGGCCCTCAGCGAGGGGCGAGGGGCTC	1800
Db	1782	CATAGTTATACCAACAGTTCCTCTCAATGACACTCAAGCCCTCAGCGAGGGGCGATGGGGTCTC	1841
QY	1801	TCATTAAGGGTCATGGTGGGCTGTCACAGACCTTAATAATGGGCAATTTGGCAAACTGCCCA	1866
Db	1842	TCACGTGGGGTCTCTGCGGTCTCACGGGACCTTAATAATGGGCAATTTGGCAAACTGCCCA	1901
QY	1861	GTTATATATGAATCTCTCCCACTAGCAAGATGGGAAGCTTGAGACTCAAAAGACTGTTTT	1920
Db	1902	GTTATATATGAATCTCTCCCACTAGCAAGATGGGAAGCTTGAGACTCAAAAGACTGTTTT	1961
QY	1921	GGACTATATGGGGAGCCCTCTGAAGGTACAACTGGACAAGCAGAGCACTGCCATCT	1980
Db	1962	GGACTTATATGGGGAGCCCTCGAAGGTACAACTGGACAAGCAGAGCTGCCATCT	2021
QY	1981	GGAGAGCAAAAGGAACAAATGACCCCACTCCCGCGGCTGAGCAGTACAGAGACT	2040
Db	2022	GAAAGAAACAAAAGGGGCCCAATGATTCAGCATCCCAAGCGGCGAGCGGGGCAAGGACT	2081
QY	2041	GACGGGAGACACAGATGTCATGACAGCAAAAGGGCAGAACCAAACTCCTGAGCTGCTGAC	2100
Db	2082	GAGGGACACACCCGCTGTCATGACAGCAAAAGGGCAGAACCAAACTCCTGAGCTGCTGAC	2141
QY	2101	ACCAATCTGATCGATGAGAGCCCTGCGCCCTTAGCCAGCTCTTTGTGCGATACAAACAA	2160
Db	2142	ACCAAGTCCGACAGATGAGAGCCCTTACCCCTTGCCAGCTCTTTGTGCGGACAAACAAAG	2201
QY	2161	GACTCCACAGGTACTTGCTGGTCTGGGTCTACACATGGAACCTCGCTCAAGAGAAAG	2220
Db	2202	GACTCCACAGGAGACTTGCTGGGCTCGGGCTTCACCACTAGGACACTCGCTCAAGAGAAAG	2261
QY	2221	CATAAATTTTTGACACAGCTCTTGACAGAGACAGTATCCCTGTGGACTTTGGCCAAAGTTA	2280
Db	2262	CATAAATTTTTGACACAGCTCTTGACAGAGACAGATTTCCCTGTGGACTTTGGCCAAAGCTG	2321
QY	2281	ACAGCAGAAGCCACAGGCCAAAGACTTGAGCCAGAGATCCAGCAGCAGCAAGCTCTTGATCA	2340
Db	2322	ACAGCAGAAGCCACAGGGCAAAAGCTTGAGCCAGAGATCCAGCAGCAGCAAGCTCTTGAGTGG	2381
QY	2341	GAACTGACTATTTAAACAAGAGCCGGTGTAGCCCCAACAAGAAAGAGATGCACTACTTGGC	2400
Db	2382	GAACTGACTGTGCAACAGGAGCCAGCGAGCCCCAAACAAAGAGAAATGCACTACTTGGCGC	2441
QY	2401	TATTTGCTAGATTAAGATGATACAAAAGATTTGGTTTACGAAATTAATTAACCCCAAACTT	2460
Db	2442	TATTTGCTCGCAAGATGATACAAAAGATTTGGTTTACCGGAAATTAACCCCAAACTC	2501
QY	2461	GAGAGACTGGAACAATGAAGATCTCGCCAGTATACCAAAATTAATGACATGAAACT	2520

Db 2502 GAGCAGCTGACAGTAAAGACATCTCGCAGTAAACAAATTAATTCATGAAGAACT 2561  
Oy 2521 GAGAAAGGAGAGATAGCTTTAGCCTGCTGACACAGCCTGGCAGTGAAGTGAACAATTG 2580  
Db 2542 GAGAAAGGAGAGATAGCTTTAGCCTGCTGACACAGCCTGGCAGTGAAGTGAACAATTG 2621  
Oy 2581 GAGAGATTTTGGATGATTTGAGAAATAGTCAATTACCACTTTTCCACAGACGAGG 2640  
Db 2622 GAGAGATTTTGGATGATTTGAGAAATAGTCAATTACCACTTTTCCACAGACGAGG 2681  
Oy 2641 CCAGCGCCCTCTGATGATCAAGTGAACAAGCCATCATCAATGACCTCATGCAACTC 2700  
Db 2682 CCAGAGCTCTCTGATGATGATGAGAAACAGTCAATACCACTTTTCCACAGACGAGG 2741  
Oy 2701 ACAGCTGAAAGACGCCCTGTCACACCTGTTGAGACCCAGAAAACAGCACTGGAATTTC 2760  
Db 2742 ACAGCTGACAGAGTCCGCTCCACCTGCCGAGGCCAGAAAGCAGCACTGCGCATGTCA 2801  
Oy 2761 CAGAGCACTTTTAATACCCAGACGAGGCAACTGGCAGGTTAATGCCAAACAGAAAT 2820  
Db 2802 CAGAGCACTTTTAATACCCAGACGAGGCAACTGGCAGGTTAATGCCAAACAGAAAT 2861  
Oy 2821 TTACCACTTGACATCAATTTGCAAAAGCCCACTGGTGTGACCTTTCCACCAATCAGA 2880  
Db 2862 TTACCACTTGACATCAATTTGCAAAAGCCCACTGGTGTGACCTTTCCACCAATCAGA 2921  
Oy 2881 AACAGTACTCCTTACTCACTGATTAATCTCAAGCAGATGATGGTATATCAAGGATGATA 2940  
Db 2922 AACATTAACCCCTTACTCACTGATTAATCTCAAGCAGATGATGGTATATCAAGGATGATA 2981  
Oy 2941 GGAACCAAGGAAATTTTGGGAAAGAGTACAGGAAATGATTTGTAACAGTCTCTCTG 3000  
Db 2982 GGAACCAAGGAAATTTTGGGAAAGAGTACAGGAAATGATTTGTAACAGTCTCTCTG 3041  
Oy 3001 CCTACTATGCCATCTGGAATATGGGACACCGCAGAGTTCGGCTGTGAGAGTCACTGTGCT 3060  
Db 3042 CCCACATATGCTTCTGGGAAATGGGACACCGCAGAGTTCGGCTGTGAGAGTCACTGTGCT 3101  
Oy 3061 GCTACCAACAGTGCATTAACCGGCGCAGTCCAGAGATGATATTCGGAACCCAGACGCC 3120  
Db 3102 GCTACCAACAGTGCATTAACCGGCGCAGTCCAGAGATGATATTCGGAACCCAGACGCC 3161  
Oy 3121 AGCATCCCATATGAGCCGACGACAGCCTGGCCTGCAAAAGCAGACCTTCAAGTCAAGTGC 3180  
Db 3162 AGCATCCCATATGAGCCGACGACAGCCTGGCCTGCAAAAGCAGACCTTCAAGTCAAGTGC 3221  
Oy 3181 ATGAATATAGGCGCATCTGAATTTAGAGATGAACATGGGGGACCTCAAGTATAGCCAACA 3240  
Db 3222 ATGAATATAGGCGCATCTGAATTTAGAGATGAACATGGGGGACCTCAAGTATAGCCAACA 3281  
Oy 3241 CAAGCTCTCCAAATCAAGACTCCCATATGCTGCTGAAAGATCTCTGCTATTAACCAAGGCG 3300  
Db 3282 CAGGCGCTCTCGAAACCAATGCTCCCGCTGAGAGATCTCTGCTATTAACCAAGGCG 3341  
Oy 3301 TCTTTTGGCAGCAAAAGCAGCAGCATTGGCAGTCTCCAGATGACTTGGTATGCTCA 3360  
Db 3342 TCGTTTGGCAGCAAAAGCAGCAGCATTGGCAGTCTCCCGCTGAGTACTGCTGTGTCA 3401  
Oy 3361 CATCTGAGCTGATCTCCAGATGATGAGGAGGCTCTCTGAGACCAAGCTGTATCTGGCC 3420  
Db 3402 CATCTGAGCTGATCTCCAGATGATGAGGAGGCTCTCTGAGACCAAGCTGTATCTGGCC 3461  
Oy 3421 TTGCGGATTTTGGATGCTGGAGAGATGATAGAGCTTATAGAAATACCCCAACTGGTGC 3480  
Db 3462 TTGCGGATTTTGGATGCTGGAGAGATGATAGAGCTTATAGAAATACCCCAACTGGTGC 3521  
Oy 3481 AGCCAGAGCCAAAGCAGTATCCAGAAACAGTCTCAAGTCAAGATTTCCAAATCATGCTG 3540  
Db 3522 AGCCAGAGCCAAAGCAGTATCCAGAAACAGTCTCAAGTCAAGATTTCCAAATCATGCTG 3581  
Oy 3541 GAGCAGAAAGGCGCGTTTCCACAGCAGTATGATCTCAAGCACAATAATGGCCCAAGGCT 3600

Db 3582 GAGCAGAAACCCCGCTTTTCCACAGCAGTACAGCATCTCAAGCACAATAATGGCCCAAGGCT 3641  
Oy 3601 AGCTATCTCCCATGCAATATCCAAATCTTCAACACCATGAGGACAGCGGCTAGTATGCG 3660  
Db 3642 GCTATTAATCCATGCAATATCCAAATCTTCAACACCATGAGGACAGCGGCTAGTATGCG 3701  
Oy 3661 ACACCTCCGATGAGCCGACGAGCGGCTCAGAGCCACGAGGCTAGTGAAGAACGCA 3720  
Db 3702 ACACCTCCGATGAGCCGACGAGCGGCTCAGAGCCACGAGGCTAGTGAAGAACGCA 3761  
Oy 3721 AATCACTAAGACTTCAACTTACAGATGCTCCCAAGCAGCAGAGATGCGCAGCCTT 3780  
Db 3762 AATCACTAAGACTTCAACTTACAGATGCTCCCAAGCAGCAGAGATGCGCAGCCTT 3821  
Oy 3781 ATGATCAATATGCAATTTTCCAAATGTTGAACCTTGTAGAGCTTGGATACCAACA 3840  
Db 3822 ATGATCAATATGCAATTTTCCAAATGTTGAACCTTGTAGAGCTTGGATACCAACA 3881  
Oy 3841 CAGGCACTAATTAATGCAAGATGCTGGCCAGAGACAGAGGAAATCTGAACAGCAT 3900  
Db 3882 CAGGCTCTAATTAATGCAAGATGCTGGCCAGAGACAGAGGAAATCTGAACAGCAT 3941  
Oy 3901 CTTCGACAGAGCAATATGATCAGCAACAGAGTTCACAAAGAACTTTGATGATGAGA 3960  
Db 3942 CTTCGACAGAGCAATATGATCAGCAACAGAGTTCACAAAGAACTTTGATGATGAGA 3995  
Oy 3961 GGACAGAGGTTGAATATGACACCAAGCAGTGTGCTCTTATGTGTGCTGACCAACTATG 4020  
Db 3996 GGACAGAGGTTGAATATGACACCAAGCAGTGTGCTCTTATGTGTGCTGACCAACTATG 4055  
Oy 4021 AGCAACCTCTGGATTTCCCAAGGCAATGCAAGCAGATTTCCATTTCCCAATCAGGA 4080  
Db 4056 AGCAATCTCCCGGATTTCCCAAGGCAATGCAAGCAGATTTCCATTTCCCAATCAGGA 4115  
Oy 4081 ATTAAGTACAGCAACTGATCAGAGCTTTACTGGGGCTACAGTCTCCAGACCCACTATG 4140  
Db 4116 ATTAAGTACAGCAACTGATCAGAGCTTTACTGGGGCTACAGTCTCCAGACCCACTATG 4175  
Oy 4141 TCACCCGGAATGGAATATCAGAGTCCCATGATGATGCAACATCTCAAGCCCAACGAGCC 4200  
Db 4176 TCACCCGGAATGGAATATCAGAGTCCCATGATGATGCAACATCTCAAGCCCAACGAGCC 4235  
Oy 4201 TATCAGGCGCCCTCGCAATTAATGATGAGGCGGCAAGCAATGGGCGGAAACAGCATG 4260  
Db 4236 TATCAGGCGCCCTCGCAATTAATGATGAGGCGGCAAGCAATGGGCGGAAACAGCATG 4295  
Oy 4261 TTTTCCAGAGTCCCGCACAACCTTTGGGAGCAAGCAACAGCAGATGTAAGTAAC 4320  
Db 4296 TTTTCCAGAGTCCCGCACAACCTTTGGGAGCAAGCAACAGCAGATGTAAGTAAC 4355  
Oy 4321 AACATGAACATGAATGTCCATGGCGAACCAACAGAGTGGCATGAGCAGATGAAACGAG 4380  
Db 4356 AACATGAACATGAATGTCCATGGCGAACCAACAGAGTGGCATGAGCAGATGAAACGAG 4415  
Oy 4381 ATGACAGGACATGACATGACATGACATGACATGACATGACATGACATGACATGACATG 4440  
Db 4416 ATGACAGGACATGACATGACATGACATGACATGACATGACATGACATGACATGACATG 4475  
Oy 4441 ATGAGTCCGAGCAGGTTAATGATCTCTGAGAGGAGAGCAACCTGTTCCCAACGAG 4500  
Db 4476 ATGAGTCCGAGCAGGTTAATGATCTCTGAGAGGAGAGCAACCTGTTCCCAACGAG 4535  
Oy 4501 CTGCGTGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4560  
Db 4536 CTGCGTGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4595  
Oy 4561 TGTCTGAGGCACTGCTTTCAGCTGACGCGGCTCAGCTTGTCTCAAAACACTTCAAGTCT 4620  
Db 4596 TGTCTGAGGCACTGCTTTCAGCTGACGCGGCTCAGCTTGTCTCAAAACACTTCAAGTCT 4653  
Oy 4621 GGAAGCTGTGTCTATTTTGTCTCAACCAACAGTACCTGCGACGCGGTTGTCTGAG- 4675  
Db 4654 GGAAGCTGTGTCTATTTGTGT- -TGCCACCTGACATGCGCGCAGTTCCTCCAGGACATA 4711

4676 - -AGCAGACAGGCGCTGGCCCTGGTCCAGGCGGCGCTCCACTCGGCTGTG-CCAGAGAG 4732  
4712 GCAGCAGACAGCTGGGGCCCTGGGCGCCGACATAGAGCCGTGCTGGCTGNNACNG 4771  
4733 AGTGCCTCTTCTTTCAGAGCTGGAAGCTGCATCCAGACAGCTGCTGCTTCCC 4792  
4772 AAGAGTGTCTCTCCGACAGCTGCCTCCAGACCAACCGCAGCTGCTGTCCAC 4831  
4793 TGCATTCACCTTAGTGCACCTTAGATCTC 4821  
4832 TGCATTCACCTTAGTGCACCTTAGATCTC 4860

RESULT 3  
AF000582 4771 bp mRNA linear ROD 25-JUN-1997  
LOCUS Mus musculus nuclear receptor coactivator protein 2 mRNA, complete cds.  
DEFINITION AF000582  
ACCESSION AF000582  
VERSION AF000582.1 GI:2213816  
KEYWORDS house mouse.  
SOURCE Mus musculus.  
ORGANISM Mus musculus; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 4771)  
Torchia, J., Rose, D.W., Inostroza, J., Kamel, Y., Westin, S., Glass, C.K. and Rosenfeld, M.G.  
The transcriptional co-activator p/CIP binds CBP and mediates nuclear-receptor function  
Nature 387 (6634), 677-684 (1997)  
TITLE 2 (bases 1 to 4771)  
Torchia, J. and Rosenfeld, M.G.  
Direct Submission  
Submitted (21-APR-1997) Medicine, Howard Hughes Medical Institute at the University of California, San Diego, 9500 Gilman Drive, C.M.M. 345, La Jolla, CA 92093-0648, USA  
JOURNAL location/Qualifiers  
MEDLINE 1. 4771  
REFERENCE /organism="Mus musculus"  
AUTHORS /db\_xref="taxon:10090"  
TITLE 152. 4543  
JOURNAL /note="NCOA-2: nuclear receptor coactivator"  
FEATURES /product="nuclear receptor coactivator protein 2"  
source /db\_xref="GI:2213817"  
cds /translation="MSGGENTSDPSRAETRRKKECPDOLGSPKSTKRNREQENK  
YIEBELIFANNDIDNENFKDKCAILKETVQJROIKEOKAANAIDVOKSDY  
SSTGQVIEKDALGPMLEBALDFEVNLEGVVFNRYTQYLNEBELNKSDY  
SILVGDHTEFVKNLPSKVNCGSNGSVEPRSSFTFNOVMYKFLPSEEBGHSO  
EAHOKYEMOCFAVSOPKSTKEGKEDLOSLIYMHDPHKGNSLLIKLYPPRP  
RODHTFGTHHESRHEAGLGRSKKDAFNSSTHSMKSLIYHNRHHYKLGAFS  
QYRFSLSGTVLVAQTKSKLRSQTTNEPQVLSLHMHREQNVVMNDLTLGAMG  
KPLNLISSSPALALCSNPGQDMTSSINPNMGPRKMGMPMGRRGSGNMHV  
SCMOATTPGOSNATLKNPSPOSSPCMNCOASVSLPRMRMGVAGSRGIPPSQFS  
PAGNLSHPGVCSSTGNSHTYSSINLALOESEGHVSSGLASPDLMKMLNSP  
VANNPPLSKMSLSDKDFGLGEPSTKGTGCAEASCHPRKKKGNDSMPAASGD  
RAEGHSRLDSKQTKLLOLLTTKSDQMEPSLPSLSIDNKSTSLPSPGTHGTS  
LKEKKHLHRLDODSSPYDLALTEAETKELSSSSSTAPOSSEVTVKQEPSPKK  
ENALRYLKDIDKTDIGLEPEITPKLERLDSKIDPASNTKLIAMKTVKESESPDO  
PGSELNLEIIDLONSOLOPEPTRPAPGVSVDKOAIIINDLMQIADSSPVPA  
GQKALIKSOSFNNRPGQGRLLPYONLPIDITLQSPGAGPPPIRNSSESVI  
PQGMANGOMLCSQCNLGNSTGMTGISTSRSPMSGEAPSTCESTLYLLPLVP  
RTDOQR EARGNPATSIIPMGANSOLGQRMLOLQVANNIGHRESENMNGGQYVQOQAP  
PNOTAPKPEIPIIDQASFASONRPFQSGPDDILCPHRAEPPSDGGLIDLYLAL  
RNFGLIEDIRALGIPELVISOQVDAEFSOESSIMLEOKPFRPOOYASQAOAO  
GGYPMQDPHEFTMGORPNYTLRMO PREGLPRTGIVONPNOLRIHRLAOQNR  
OPLMNOISSYNNLTLRGVPTOABINOMLQORRELLNOLIROMQOYQORTL  
MRGQGLNTPSVAVPAGLPAANSNRPTRPANKQPPPPPTISQSDPDEGTATTP  
OSP LMSPRMAHTOSPMQOSQANPAVOTYDMNGMAQSGMGSNMF500SPHFGQDA

NTSMTSNNNITVSATNTGGLSSNMQNTGQMSMTSVTSVPTSGLPMSQPEVNDPAL  
RQGNLFEPNQLMDIMIKDEGDSRRYC"  
BASE COUNT 1292 a 1347 c 1228 g 904 t  
ORIGIN  
Query Match 60.1%; Score 3699.6; DB 10; Length 4771;  
Best Local Similarity 87.7%; Pred. No. 0;  
Matches 4194; Conservative 0; Mismatches 554; Indels 32; Gaps 13;  
10 AGCTGGGCTACAGCTTGGCGGGAAGGTCACGGCGGAGCGGACCTGACGCG 69  
1 AGCTGGGCTCCCGCGGCGGCGGAGGTACAGCGCGGAGCGGCGGCGGCGGCG 60  
70 GTGACGACCCGAGCCGATTTCTTGGATTGGCTACACACTTAATAGATCTTCTGCACT 129  
61 GTGACGS -CCCATTTGATTTTCTGCGATCTGGCTTCACTGATTTGGCTTCTGCACT 118  
130 GTTTACAGSCACAGTTGCTGATATGTGTTACAGATGAGTGGGATGGAGAAATACCTCT 189  
119 GTGTACAGGACAGATTGCTGATATGTGTTCAAGATGAGTGGGATGGAGAAACACCTCT 178  
190 GACCCCTCCAGGCGAGAGCAAGAAAGCGAAGAAATGCTGACCACTTGAGCCGACG 249  
179 GACCCGTCAGGCGAGAGCAAGAAAGCGAAGAAATGCTGACCACTTGAGCCGACG 238  
250 CCCAAAGAGCACTGTAACGTAATCGTAACAGGAAATTAATATAGAAAGACTT 309  
239 CCCAAAGAGCACTGTAACGTAATCGTAACAGGAAATTAATATAGAAAGACTT 298  
310 GCAAGTTGATTTTGTCAATTTTATGATATAGACAACTTTAAGTTCAAGTACGACAA 369  
299 GCGGAGCTGATCTTCGCAAACTTAAATGATATGACAACTTCAAGTCAAGTCAAGTCA 358  
370 TGTGCACTCTTAAAGAACTGTGAAGCAATTCGTGATGTCAAAGCAAGAAAGCA 429  
359 TGTGCACTCTTAAAGAACTGTGAAGCAATTCGTGATGTCAAAGCAAGAAAGCA 418  
430 GCAAGTCCCAACATAGATGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 489  
419 GCAAGTCCCAACATAGATGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 478  
490 ATGACAGAGATGCGCTGGGCGCTATGATGCTTGAAGCCCTTATGAGGCTTCTTGTGA 549  
479 ATGACAGAGATGCGCTGGGCGCTATGATGCTTGAAGCCCTTATGAGGCTTCTTGTGA 538  
550 GTGACCTGGAAGCAAGCTGTTGTTGTGTCAGAGAAATGTCACACAGTATTAAGTAT 609  
539 GTGACCTGGAAGCAAGCTGTTGTTGTGTCAGAGAAATGTCACACAGTATTAAGTAT 598  
610 AACCAAGAGAGTGAATGAACAAAGATGATATAGATGATGATGATGATGATGATGAT 669  
599 AACCAAGAGAGTGAATGAACAAAGATGATATAGATGATGATGATGATGATGATGAT 658  
670 GAATTTGTCAAAAACCTGCTGCCAAAGTCTATAGTAAATGGGGATCTTGGTCTGGGAA 729  
659 GAATTTGTCAAAAACCTGCTGCCAAAGTCTATAGTAAATGGGGATCTTGGTCTGGGAA 718  
730 CCTCCGAGGGGGAACGACCATCTTCAATTTGTCGATGTCGATGTCGATGTCGATGTC 789  
719 CCTCCGAGGGGGAACGACCATCTTCAATTTGTCGATGTCGATGTCGATGTCGATGTC 778  
790 TCAGAAAGAGAGGTCATGATTAACAGAGAGTCAATCAGAAATATGAAATATGACAGTGC 849  
779 TCAGAAAGAGAGGTCATGATTAACAGAGAGTCAATCAGAAATATGAAATATGACAGTGC 838  
850 TTGCTGCTCTCTACCAAGTCCATCAAGAAAGAGAGAGAGATTTGCACTGCTGCTTG 909  
839 TTGCTGCTCTCTACCAAGTCCATCAAGAAAGAGAGAGAGATTTGCACTGCTGCTTG 898  
910 ATTGCGTGCAGAAAGATTTCCATGAAGAAAGAGAGAGATTTGCTCCCTCATGAGAAAT 969  
899 A-TTGTGTGCACGAGA-TCCCATGAAGAAAGAGAGAGATTTGCTCCCTCATGAGAAAT 936





```

Db 3117 TCCCCATGGAGGAGCCACAGCAGCTTGCGCAAGACAGATGCTTCAAGTCACAGTCATGA 3176
QY 3185 ATATAGGCGCATGATGATTAAGATGAACATGGGGGACCTCAGTATAGCCACACAG 3244
Db 3177 ACATAGGCGCTTGTGATGATGAATGAATGAGAGGACCTCAGTATTAATCAACACAG 3236
QY 3245 CTCCTCCAAATGACAGTGGCCCTGAAAGCATCTGCTTATAGACAGGCGCTT 3304
Db 3237 CCCCTCGAACCAGAACCTGGCCCGCTGAGAGCATCTGCTTATAGACAGGACATCG 3296
QY 3305 TTGCCAGCCAAAACAGGACGACCTTTGGCAGTTCTCAGATGACTTGTATGTCACATC 3364
Db 3297 TTGCCAGCCAGAAACAGGACGACCTTGGCAGCTCCCTGATGACCTGTGTGTCACATC 3356
QY 3365 CTGCACCTGATGCTCGAGATGATGAGGAGCTCTCTGGACAGCTGATCTGGCTTGC 3424
Db 3357 CTGCACAGAGTGTGCGAGGATGAGGGGCGCTCTTCTTGACAGCTGTATCTGGCTTGC 3416
QY 3425 GGAATTTTATGGCTGAGAGATGATGATGAGCTTATAGAAATACCGAATGGTGAGCC 3484
Db 3417 GGAATCTTCAATGGCTTGAAGATTTGATGATGAGCTCTGGGATATACGAATGCTGAGCC 3476
QY 3485 AGAGCCAGCAGTATGATCCAGAACATTTCTCAAGTACGATTTCAACATCATCTGAGCC 3544
Db 3477 AGAGCCAGCAGTATGATGAGAGAGCTTCTCAAGTACGAGCTCAGCATATATGCTGAGC 3536
QY 3545 AGAAGGCGCGCTTTTCCACAGCAGTATGATCTCAGGACAAATAGGCGCAGGAGCT 3604
Db 3537 AGAAGCGCGCGCTTTTCCACAGCAGTATGATCTCAGGACAAATAGGCGCAGGAGCT 3596
QY 3605 ATTTCCATGACAGATTCACAACTTTCACACATGAGGACAGCGGCTAGTTATGCCACAC 3664
Db 3597 ATATCCCATGACAGATTCACAACTTTCACACATGAGGACAGCGGCTAGTTATGCCACAC 3656
QY 3665 TCCGTATGACAGCCAGACCGGCGCTCAGGCGCCAGGCGCTAGTGTGAGAACACCAATC 3724
Db 3657 TCCGTATGACAGCCAGACCGGCGCTCAGGCGCCAGGCGCTAGTGTGAGAACACCAATC 3716
QY 3725 AACTAGACCTTCAACTTCAGCATGCGCTCCAAAGCAGACAGAGATGCGCAGCATTATGA 3784
Db 3717 AACTAGACCTTCAACTTCAGCATGCGCTCCAAAGCAGACAGAGATGCGCAGCATTATGA 3776
QY 3785 ATCAAAATCAGCAATGTTCCAAATGTAATGTAATGTAATGTAATGTAATGTAATGTA 3844
Db 3777 ATCAAAATCAGCAATGTTCCAAATGTAATGTAATGTAATGTAATGTAATGTAATGTA 3836
QY 3845 CAGCTTATATGACAGATGCTGGCCAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3904
Db 3837 CAGCTTATATGACAGATGCTGGCCAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3896
QY 3905 GACAGAGACAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3964
Db 3897 GCGAGAGACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3950
QY 3965 AAGGGTGAATATGACACAGCAGTGTGCTTCTAGTGTATGACAGCAATATGAGCA 4024
Db 3951 AAGGGTGAATATGACACAGCAGTGTGCTTCTAGTGTATGACAGCAATATGAGCA 4010
QY 4025 ACCCTGGAGTTCGCCAGGCAATGACAGAGATTTCCATTTCTCCAAACTAGGAAATTA 4084
Db 4011 ATCCCGGAGTTCGCCAGGCAATGACAGAGATTTCCATTTCTCCAAACTAGGAAATTA 4070
QY 4085 GTCAGACACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4144
Db 4071 GTCAGACACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4130
QY 4145 CCCGAATGACATACACAGAGTCCCATGATGACAGAGTCTAGGCGCAACGACCTATC 4204
Db 4131 CCCGATGACATACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4190
QY 4205 AGGCGCCCTCCGATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4264

```

```

Db 4191 AGCCACCTTCAGACATGAATGATGATGATGATGATGATGATGATGATGATGATGATG 4250
QY 4265 CCCAGCAGTCCCGACACACATTTGGGAGAGCAAGCAACACAGCATGATGATGATGATGATGATG 4324
Db 4251 CACAGCAGTCCCGACACACATTTGGGAGAGCAAGCAACACAGCATGATGATGATGATGATGATG 4310
QY 4325 TGAACATCAATGTGTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4384
Db 4311 TGAACATCAATGTGTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4370
QY 4385 CAGGACAGTACAGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4444
Db 4371 CAGGACAGTACAGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4430
QY 4445 GTCCGAGAGGATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4504
Db 4431 GTCCGAGAGGATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4490
QY 4505 CTGGAATGATATGATTAAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4564
Db 4491 TTGGAATGATATGATTAAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4550
QY 4565 GAAGCCAGTGTCTTCTCAGCTGACCGGCTCAGTTGCTCAAAACCTTCCAGTCTGAGAG 4624
Db 4551 GAACCTGTGATGATTTTTCACACCCACTGGGCTT -ACAAACATTTACAGCTGTGAG 4608
QY 4625 AGCTGTCTATTTTGTTCACACCCACTGGGCTT -ACAAACATTTACAGCTGTGAG 4676
Db 4609 AGCTGTCTCTTCTTGTGT -TGCCACCTGACATGCGGCGAGTCTCCACAGACATGAG 4666
QY 4677 -GCAGACAGCGCTGGCCCTGTTCCAGAGGTGAGGCTCAGTCTGAGGAGAGAGAGAGAG 4733
Db 4667 CAGACAGTGGGCGCTGGGCGGCGGACAGCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4726
QY 4734 GCTGCTCTTCTTCTTACAGCTGAGAGCTGAGAGCTGAGAGCTGAGAGCTGAGAGCTGAG 4773
Db 4727 GTTGCTCTCTCC -GACAGCTGAGCTGCTGCTCCAGAC 4764

RESULT 4
AF136943 4398 bp mRNA linear ROD 12-APR-1999
LOCUS AF136943 Rattus norvegicus transcriptional intermediary factor 2 (Tif2)
DEFINITION Rattus norvegicus transcriptional intermediary factor 2 (Tif2)
ACCESSION AF136943
VERSION AF136943.1 GI:4581052
KEYWORDS Norway rat.
SOURCE Rattus norvegicus
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 4398)
Leers,J., Treuter,E. and Gustafsson,J.-A.
Mechanistic principles in NR box-dependent interaction between
nuclear hormone receptors and the coactivator Tif2
Mol. Cell. Biol. 18 (10), 6001-6013 (1998)
98414616
JOURNAL 2 (bases 1 to 4398)
Leers,J., Treuter,E. and Gustafsson,J.-A.
Direct Submission
Submitted (24-MAR-1999) Department of Biosciences, Karolinska
Institute, Haelsovaegen 7, Huddinge 14157, Sweden
LOCATION/Qualifiers
1..4398
FEATURES
source
1..4398
/gene="Tif2"
/db_xref="taxon:10116"
1..4398
/gene="Tif2"
/function="nuclear hormone receptor coactivator"
/codon_start=1

```



```

/product="transcriptional intermediary factor 2"
/protein_id="AAD24587.1"
/db_xref="GI:4581053"
/translation="MSGGENTSDPSRAETRRKRECPDOLGSPSRSTENKRENDENK
YIEELAEIIFANFDIDNFNPKPKCALIKETVQIRIKDEKRAAANIDEVOKSDY
SSTOGVYDKDALGPMLEALDGFEEVNLGNVPESENTQYLRNOBELMKSIV
SILHVGHTEFEVKNLIPKSNVNGSMTEPRRNSHPNLMYKPLDSEEGNDNO
EAHQKYTEMOCFAVQSPKSIKEEGDKOSCLIVARVPKMERKPLPSESSEFTDOL
OGKITFLDITSTVRAMKPGMEDIVRCLQKHTHDEGSLYAKRHHEVYRLOGIAGS
OIVREPLSDGLTVAQFKSKLRISQTPMEPLVLSIHMLHEQVNCVNNPLDGLAMC
KPLPSHSSSPARAMCGANGODVALCSNNFPMNCRREDHSMRFGSGGNNHV
SGMOATTPQGSNVALKMSPSOSSGLNPPSSVLSRHRMSRQVAVSPVPQFS
PAGSLHPAGVCSSTGNHSYSTNSLNLQALSRBHGVSPLASPLDKMGNQNSP
VNMNPPLSKMSGLSDKDFGLYGESEGTGQAOASPEOKRPNDSMPQASD
RAEGHSLRHSKGGOTKILQTLTKSDQMEPLDLSLDTKDQSLDPSGSGTGTG
LKEKHKLILRLDQSSPVDLAKLTAETGELNDESGTAPGSEVTVYKQOAPSPKK
ENALLRVLDDKDPDKGLPEITRKLELDSKTPASNTKILANKTYEEVSEFSDQ
PGEELDNLEELDLQNSQLPOLPEPDRPGAPTSVDQALINDLMQTLADSPVTPY
GAQRALALMSQSTFNNRPGQLRLPQNLPLDITLOSPGAGFPPIRNSPVSIV
POPMGNOGMLGSGNLSGMSSTGSMISGEMAPQSGEMAPQSGEMAPQSGEMAP
NRPIOGGIMRPTASIPMRANSQGOROMLOPOYMNIGSELEMMMGPOYNQOQAP
NOTAPWPESTILPIDOASGSONRHRPSSPDLLCPHRAESPDEGALLDQILALR
NFDLELEIDRALGIPELVYSQAVDPEQFSQESSMILEKPPVPPOOYASOTMAQ
SYNPMODENFTMGORPNYTTLRMQPRGLEPRTGIVQNPOLRLQLOHRLQAOQNRQ
PLMNQISGVSNVNLTLRGPVPTQAPINAMQLOBORELINQHLRQOMHQOQOQRT
LMMQOGILNMTPSMVAPTGLPAASNPRIPOANAOQEPFPNYGISOODPDEGTGAT
POSLMSPRMAHTQSPMMQOSQANPAYOPADIDNGMOSGMSMSFQOSDPHFGQO
ANTSMYNNNNINYSMAINTAGLSNNMNMQOMMTSVTSVPTSGLSMGPEQVNDPA
LRGSSLETTNOLPMDMIKQEDGDSKTC"
BASE COUNT      1233 a      1232 c      1125 g      808 t
ORIGIN

```

```

Query Match      58.4%: Score 3593.2; DB 10; Length 4398;
Best Local Similarity 88.7%: Pred. No. 0;
Matches 3902; Conservative 0; Mismatches 493; Indels 3; Gaps 1;

```

Qy	163	ATGAGTGGATGGAGAAATACCTCTGACCCCTCCAGGGCAGAGACAAGAGGCGAAG	222
Db	1	ATGAGTGGATGGAGAAATACCTCTGACCCCTCCAGGGCAGAGACAAGAGGCGAAG	60
Qy	223	GAATGCTCTGACCACTTGACCCAGCCCCAAAGAAACACTGAAAAAAGTAACTGTGA	282
Db	61	GAATGCTCCGACACCTGACCCAGCCCCAAAGAAACACTGAAAAAAGTAACTGTGA	120
Qy	283	CAGGAAATTAATATATAGAAACTTGCAGAGTTGATTTTGCAGAAATTTAATATATA	342
Db	121	CAGGAAATTAATATATAGAAACTTGCAGAGTTGATTTTGCAGAAATTTAATATATA	180
Qy	343	GACAACTTAACTCAAACTGACAAATGTGCATCTTAAAGAAAGTGTGAAGCAAT	402
Db	181	GACAACTTCAAACTGACAAATGTGCATCTTAAAGAAAGTGTGAAGCAAT	240
Qy	403	CGTCAGATCAAGAAGCAAGAGAAACAGCAGCTGCAACATATAGATGAAGTGCAGAGTCA	462
Db	241	CGCCAGATCAAGAAGCAAGAGAAACAGCAGCTGCAACATATAGATGAAGTGCAGAGTCA	300
Qy	463	GATGATCTCTACAGGGCAGGGGTGCATTCGACAGAGATGGCGTGGGCTATGATGCTT	522
Db	301	GATGATCTCTACAGGGCAGGGGTGCATTCGACAGAGATGGCGTATGATGATGCTT	360
Qy	523	GAGGCCCTTGATGGTCTTCTTTGTAGTGAACCTGGAAGGCAAGTGTGTGTGTCA	582
Db	361	GAGGCCCTTGATGGTCTTCTTTGTAGTGAACCTGGAAGGCAAGTGTGTGTGTCA	420
Qy	583	GAGAAATGACACAGTATCTTAAGGTATTAACAAGAGAGCTGATGACAAAAGTGTATAT	642
Db	421	GAGAAATGACACAGTATCTTAAGGTATTAACAAGAGAGCTGATGACAAAAGTGTATAC	480
Qy	643	AGCATCTTGATGGGGACACAGGAATTTGCAAAAACCTCTGCAAGAGCTATA	702
Db	481	AGCATCTTGATGGGGACACAGGAATTTGCAAGAACCTCTGCAAGAGCTATA	540

Qy	703	GTAATGGGGGATCTTGCTGTGCGAACCCTCGAGGCGGAGACAGCCATACCTTCAATGCT	762
Db	541	GGAAATGGGGGATCTTGCTGTGCGAACCCTCGAGGCGGAGACAGCCATACCTTCAATGCT	600
Qy	763	CGATGCTGTGTAACCTTACCTGTAGAGAGAGAGAGGTCATGATTAACAGAGAGCT	822
Db	601	CGATGCTGTGTAACCTTACCTGTAGAGAGAGAGAGGTCATGATTAACAGAGAGCT	660
Qy	823	CATCGAAATTAAGAACTATGACAGTCTCGCTCTCTCAACCAAGTCCATCAAGAA	882
Db	661	CATCGAAATTAAGAACTATGACAGTCTCGCTCTCTCAACCAAGTCCATCAAGAA	720
Qy	883	GAGAGAGAAATGATTCAGTCTGCTTGTATTCCTGTGCAAGAGAGTTCCTATGAAGAA	942
Db	721	GAGGGGGAAGATTAAGAGATGCTGCTTGTATTCCTGTGCAAGAGAGTTCCTATGAAGAA	780
Qy	943	AGACAGTCTTCTCCATCAGAAAGTTTACTCTGCGCAGAGATCTCCAGGCAAGATC	1002
Db	781	AGGCGAGCCCTTCCCTCATCAGAGAGCTTACACCCGCGCAGAGACTCCAGGCAAGATC	840
Qy	1003	AGTCTCTGATTAACAGCAGCAGCAGAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAG	1062
Db	841	ACTTTTCTGACACATGACACATGACAGAGATGACAGAGATGACAGAGATGACAG	900
Qy	1063	AGAAGGTGATTCAGAAAGTTCATGCGCAGCATGGAAGAGAAATCTGTCTATGCTAAG	1122
Db	901	AGAAGGTGATTCAGAAAGTTCATGCGCAGCATGGAAGAGAAATCTGTCTATGCTAAG	960
Qy	1123	AGGATCATCATGAAAGTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	1182
Db	961	AGGATCATCATGAAAGTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	1020
Qy	1183	TTGTCTGATGAGCAGCTTGTCTGACAAAGAGAGAGAGAGAGAGAGAGAGAGAG	1242
Db	1021	TTGTCTGATGAGCAGCTTGTCTGACAAAGAGAGAGAGAGAGAGAGAGAGAGAG	1080
Qy	1243	ACTAATGAACCTCACTTCTAATATCTTATCATATGCTTCAAGAGAGAGAGAGAG	1302
Db	1081	ACGAATGAGCCTGACCTTGTGATGATGATGATGATGATGATGATGATGATGATG	1140
Qy	1303	GTGATGATTCGAGATCTGACCTGAGCAAAAGAGAGAGAGAGAGAGAGAGAGAG	1362
Db	1141	GTAAATGATTCGAGATCTGACCTGAGCAAAAGAGAGAGAGAGAGAGAGAGAGAG	1200
Qy	1363	AACAGCCCTGCGCCATCAGAGCCCTGTGCAAGTGGGAACCCAGAGTCAAGACATGACCTCAAGT	1422
Db	1201	AACAGCCCTGCGCCATCAGAGCCCTGTGCAAGTGGGAACCCAGAGTCAAGACATGACCTCAAGT	1260
Qy	1423	AGCAATATTAATTTTCCATTAATGAGCCCAAGAAAGAAATGGGATGGCCATGGGCAAG	1482
Db	1261	AGCAATATTAATTTTCCATTAATGAGCCCAAGAAAGAAATGGGATGGCCATGGGCAAG	1320
Qy	1483	TTTGTGTTCTGGGGGAGATGATGATGATGATGATGATGATGATGATGATGATG	1542
Db	1321	TTTGTGTTCTGGGGGAGATGATGATGATGATGATGATGATGATGATGATGATG	1380
Qy	1543	AGTAATATGCACTCAAAATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	1602
Db	1381	AGTAATATGCACTCAAAATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	1440
Qy	1603	CAGCCACCTCCATCTTTCACCAAGGATGCAATGAGAGAGAGAGAGAGAGAGAG	1662
Db	1441	CAACCAAGTCTGTCTCTCCCAAGGATGCAATGAGAGAGAGAGAGAGAGAGAG	1500
Qy	1663	CGAATCCACCAAGTCTGTCTCTCCCAAGGATGCAATGAGAGAGAGAGAGAGAG	1722
Db	1501	CGGCTTCCACCAAGTCTGTCTCTCCCAAGGATGCAATGAGAGAGAGAGAGAG	1560
Qy	1723	AGCAGCAGAGAAATAGCATATGATTAACCAACAGCAGCAGCAGCAGCAGCAG	1782
Db	1561	AGCAGCAGAGAAATAGCATATGATTAACCAACAGCAGCAGCAGCAGCAGCAG	1620
Qy	1783	AGCAGGAGGAGAGGAGTCTCATATAGGTCATATGAGTCTTACAGAGAGAGAG	1842

```

1621 AGGAGGGCCACGGGGCTGTTAGGGCCATCGTTGGCTTCCTCCGGACCTTAATAATGGG 1680
1843 AATTGGAAAACTCCCGAGTTAATATGATCTCCCGACCTCAGCAGATGGGAAGCTTG 1902
1681 AATTGGAAAACTCCCGAGTTAATATGATCTTCCTCCCGACCTCAGCAGATGGGAAGCTTG 1740
1903 GACTCAAAAGACTGTTTGGACTATATATGAGAGCCCTCTGAAAGGTACACTGCAACAGCA 1962
1741 GACTCCAAAGACTGTTTGGACTTATGAGGAGCCATCAGAAAGGTACACTGCAACAGCA 1800
1963 GAGAGCGCTGCCATCCTTGAGAGCAAAAGAAAACAATGACCCCACTGCCGCCGCCGCC 2022
1801 CAGGCGAGCTGCCATCCTTGAGAGCAAGAGAGGCCCAATATCTCAGCATGCCCAAGGCG 1860
2023 GTGAGCAGTGAAGAGCTGACGGGCGAGAGAGAGCTGATGACAGCAAAAGGCGAGCAAA 2082
1861 GCCAGGAGAGACAGGGCTTGAGGAGACAGCCGGCTGCATGAGAGCAAAAGGCGAAACCAA 1920
2083 CTCCTGCAGCTGCTGACCCACCAATCTGATCAGATGAGAGCCCTGCGCCCTTAGCCAGCT 2142
1921 CTCCTGCAGCTGCTGACCCACCAAGTCCGACAGATGAGAGCTTCACCTTGCCAGCTCT 1980
2143 TTGTGCAATCAAAACAAGACTCCACAGTAGCTGTGCTGCTGCTGCTGCTGCTGCTGCTG 2202
1981 TTGTGCAACAACAAGACTCCACAGGAGCTTGCGAGGGCTTGCGTCCAGCATGGC 2040
2203 ACCTCGCTCAGAGAGAGACTTAAATTTTGCACAGACTTTGCGAGCAGCAGTTCCTCT 2262
2041 ACCTCGCTCAGAGAGAGACTTAAATTTTGCACAGACTTTGCGAGCAGCAGTTCCTCT 2100
2263 GTGAGACTTGCCCAAGTTTAACAGCAGAAAGCCACAGAGCAAGCTGAGCCAGAGTCCAGC 2322
2101 GTGAGACTTGCCCAAGTTTAACAGCAGAAAGCCACAGAGCAAGCTGAGTCAAGATCCAGT 2160
2323 AGCAGAGCTCTGGATCAGAGTAGATTTAACAAGAGAGCGCGTGGAGCCCAAGAAACAA 2382
2161 GGCAGGCTCTGGATCAGAGTAGATTTAACAAGAGAGCGCGTGGAGCCCAAGAAACAA 2220
2383 GAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2442
2221 GAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2280
2443 GAAATTAACCCCAACTTGAGAGACTGAGACATGTAAGACAGATCCTGCCAGTACACAAA 2502
2281 GAAATTAACCCCAACTTGAGAGACTGAGACATGTAAGACAGATCCTGCCAGTACACAAA 2340
2503 TTAATGCAATGAAAACTGAGAGAGAGAGATGAGCTTGAGCCTGGTGAACAGCCATATC 2562
2341 CTATATGCTGATGAAACTGTAAGAGAGAGAGATGAGCTTGAGCCTGGTGAACAGCCATATC 2400
2563 AGTAGCTGAGCAACTTGAGAGAGATTTGATGATTTGAGAGATGATGATGATGATGATGATG 2622
2401 AGGAGCTGAGCAACTTGAGAGAGATTTGATGATTTGAGAGATGATGATGATGATGATGATG 2460
2623 CTTTTCCAGACACAGAGGCGAGCGCCCTGCTGATGATGATGATGATGATGATGATGATGATG 2682
2461 CTTTTCCAGACACAGAGGCGAGAGCTCTGATGATGATGATGATGATGATGATGATGATGATG 2520
2683 AATGAGCTCTATGCAACTGACAGTGAAGAAAGAGCCCTGTACACTGTGTGAGAGCCGAGAA 2742
2521 AATGAGCTCTATGCAACTGACAGTGAAGAAAGAGCCCTGTACACTGTGTGAGAGCCGAGAA 2580
2743 ACAGCACTGGGAATTTGACAGAGCACTTTAATTAACCCAGAGACAGAGCAACTGGGAGG 2802
2581 GCAGCACTGGGAATTTGACAGAGCACTTTAATTAACCCAGAGACAGAGCAACTGGGAGG 2640
2803 TTATTTGCCAAACAGAAATTTACCACTTGACATCAGATTGCAAAAGCCCACTGTGCTGCA 2862
2641 TTATTTGCCAAACAGAAATTTACCACTTGACATCAGATTGCAAAAGCCCACTGTGCTGCA 2700
2863 CCTTTCCACCAATCAGAAACAGTAGTCCCTACTCAGTGTACCTCAGCCAGGAATGATG 2922
|||||
2701 CCTTTCCACCAATCAGAAACAGTAGTCCCTACTCAGTGTACCTCAGCCAGGAATGATG 2760
2923 GGTAAATCAAGGATGATGAGAAACCAAGAAATTTAGGAAACAGTAGACAGGAATGAT 2982
2761 GGTAAACCAAGGATGATGAGAAACCAAGAAATTTAGGAAACAGTAGGAAACATATGACAGGAATGAT 2820
2983 GGTAAACAGTCTTCTGAGCTACTATGTCATCTGAGAAATGGCCACCCAGAGTGTGGCT 3042
2821 GGCAGTGTGACTTCCCGTCCAGCATGCTCTTCTGAGGAAATGGCCACCCAGAGTGTGCAAGCT 2880
3043 GTGAGAGTACCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3102
2881 GTGAGAGTACCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2940
3103 ATTGGAACCCAGCAGCAGCATCCCATATGAGAGCCCGAGCAGCCAGCTGTGCGCAAGACAG 3162
2941 ATTGGAACCCAGCAGCAGCATCCCATATGAGAGCCCGAGCAGCCAGCTGTGCGCAAGACAG 3000
3163 ACGCTTCACTGCTGAGTCAATGAATATATGAGGCGCATCTGCAATTTAGAGATGAACATGGGGGA 3222
3001 ATGCTGAGCCTCAGGTCATGAACATGAGGCGCTTCTGATGAGATGAACATGAGGAGCG 3060
3223 CCTCAGTATGCGCAACAGACAGCTCTCCCAATCAGACTGCGCCCATGCGCTGAAAGCATC 3282
3061 CCTCAGTATATCAACAGACAGGCGCTTCCCAACCAACTCCCTCGGCGCGAGAGCATC 3120
3283 CTCGCTATAGACAGGCGCTCTTTTGGCAGGCCCAAAACAGCAGCAGCATTTTGGCAGTTCGA 3342
3121 CTCGCTATATCAACAGAGCATTTTTCGGCAGCCAGAAAGGAGCATCCCTTGGCAGCTCCCC 3180
3343 GATGACTTGTATGTGTCACATCTCTGACGTGAGTCTCCGAGTGTAGGAGGAGCTCTCTG 3402
3181 GAGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3240
3403 GACAGCTGATATGAGGCGCTTGGCAATTTTGAATGGCTGAGAGATGATGATGAGAGCTTGA 3462
3241 GACAGCTGATATGAGGCGCTTGGCAATTTTGAATGGCTGAGAGATGATGATGAGAGCTTGA 3300
3463 GGAATATCCCACTGCTGACAGCCAGAGGCGCAAGAGTAGATCAGAAACAGTTCCTCAAGTCAG 3522
3301 GGGATATCAAGAACTGTCGACAGCCAGAGGCGCAAGAGTAGATCAGAAACAGTTCCTCAAGTCAG 3360
3523 GATTTCAACATCATGCTGGAGCAGAAAGGCGCGCTTTTCCACAGCAGTATGCTATCAG 3582
3361 GAGTCCAGCATGATGCTGGAGCAGAAAGGCGCGCTTTTCCACAGCAGTATGCTATCAG 3420
3583 GCACAAATGGCCCGAGGCTACTATCTCCCATGCAAGATCAAACTTCCACACATGAGGA 3642
3421 ACACAAATGGCCCGAGGAGACTATATCCATGCAAGATCAAACTTCCACACATGAGGA 3480
3643 CAGGCGCTGATATGCTGACACTCCGATATGACAGCCAGAGCCGAGGCGCTCAGGCGCAGGCG 3702
3481 CAGGCGCTGATATGCTGACACTCCGATATGACAGCCAGAGCCGAGGCGCTCAGGCGCAGGCG 3540
3703 CTAGTGCAGAACCAAGCCAAATCAACTAGACTTCACTTACATGAGTGGCTCTCAAGCAGAG 3762
3541 ATTGTGAGAGAACCAAGCCAAATCAACTAGACTTCACTTCAAGATGCTCTCAAGCAGAG 3600
3763 CAGAAATGCGAGCAGCTTATGAAATCAAAATCAGAAATTTTCCAAATGAGTGAATGAGTCTG 3822
3601 CAGAAATGCGAGCAGCTTATGAAATCAAAATCAGAAATTTTCCAAATGAGTGAATGAGTCTG 3660
3823 AGGCTTGAATGATGACAGCAGCTTATGAAATGAGAGATGATGAGTGGCCCGAGAGCAGAGG 3882
3661 AGGCTTGAATGATGACAGCAGCTTATGAAATGAGAGATGATGAGTGGCCCGAGAGCAGAGG 3720
3883 GAAATCTGAGACAGACTTCTTGAAGAGACAGCAAAATGATCAGAGCAAGATTGACAGAA 3942
3721 GAAATCTTAAACAGCATCTCCGAGAGAGACAGATGATGATGATGATGATGATGATGATGATG 3780
3943 CGAATCTTGAATGATGAGAGACAGAGGTTGAATATGACACCAAGCATGTGGCTCTGATG 4002
3781 CGAATCTTGAATGATGAGAGACAGAGGTTGAATATGACACCAAGCATGTGGCTCTGATG 3840
|||||
```

OY	4003	GSTATGCCAGCACTATGACGAACCCCTGGATTTCGCCAGCAAATGCAOAGCATTTTCCA	4062
Db	3841	GGTTTACCAGCAGGCATGAGCAACCCCGGGATTCGCCAAGCCAAATGCCAGTTTCCA	3900
OY	4063	TTTTCTCCAAACTACGGAATAAGTAGCAGAACCTGTATCCAGGCTTAACGGGGCTACGACT	4122
Db	3901	TTTTCTTCCGAACATGAGGAATAGTGAGCAACCTGTACTCTGGTTTACTGAGACTCAAACT	3960
OY	4123	CCCCAGAGCCCACCTTATGTTCACCCCGAATGGCACATACACAAGAGTCCCATGTATGCACAG	4182
Db	3961	CCCCAGAGTCCCTTAATGTCTCCCGGATGGGACATATGCAAGAGTCCCATGTATGCAGCAG	4020
OY	4183	TCTCAGGCCAACCCAGGCTTATCAGGCCCCCTCCGACATPAATGTGATGGCGCAGGGGAAC	4242
Db	4021	TCTCAGGCCAACCCAGGCTTACCAAGCCCCCTCAGACATTAAGGATGGGCAACAGGGGAAGC	4080
OY	4243	ATGGCGGAAAAACAGCATGTTTTCCAGAGAGTCCCAACACACTTTGGGCGAGCAACCAAC	4302
Db	4081	ATGGCGGAAAAACAGCATTTTCTCACACAAGTCCCAACACACTTTGGGCGACMAACCGAAC	4140
OY	4303	ACCAAGCATGTACAGTAAACATGAACATGAATGTCTCCATGGCGGCAACACAGATGGC	4362
Db	4141	ACCAAGCATGTAAATTAACAACATGAACATGAACATGTCATATGGCAACCAACAGGCTGGC	4200
OY	4363	ATGAGCAGCATGAACACCAATGACAGGACAGATGACATGACATGACCTCAGTACCTCGCTGCT	4422
Db	4201	TTGAGCAACATGAACCAATGACAGGCGCAGATGACATGACATGACCTCAGTACCTCGTGCCT	4260
OY	4423	ACGTCAGGAGCTGTCCTCCATGGGTCCGAGCAGGTAAATGATCCGTCTGTGAGGGAGGC	4482
Db	4261	ACATTCAGGGCTGTCTCCTCATGGGTCTTAACAGGTCATGACCTCGTCTGAGGGGAAGC	4320
OY	4483	AACCTGTT---CCCAAAACCACTGCGCTGGAATGATGATTAATTAAGCAGAGGAGACACA	4539
Db	4321	AGCCTTTTCACACACAAACCACTGCGTGAATGAGCATGATCAACAGCAGAGGAGATGGG	4380
OY	4540	ACAGGAAATATMTGCTGA	4557
Db	4381	TTCTGGAAATATCTGCTGA	4398
RESULT	5		
LOCUS	XLA243119	4871 bp	mRNA linear VRT 07-APR-2000
DEFINITION	Xenopus laevis mRNA for transcription intermediary factor 2 (tlf2 gene).		
ACCESSION	AJ243119		
VERSION	AJ243119.1	GI:5123920	
KEYWORDS	TLF2 gene; transcription intermediary factor 2.		
SOURCE	African clawed frog.		
ORGANISM	Xenopus laevis		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae; Xenopodinae; Xenopus.		
AUTHORS	1 (bases 1 to 4871) de la Calle-Mustienes, F. and Gomez-Skarmeta, J.L.		
TITLE	xrlf2, a Xenopus homologue of the human transcription intermediary factor, is required for a nuclear receptor pathway that also interacts with CBP to suppress Brachyury and Xmyd		
JOURNAL	Mech. Dev. 91 (1-2), 119-129 (2000)		
MEDLINE	20171035		
REFERENCE	2 (bases 1 to 4871) Gomez-Skarmeta, J.L.		
AUTHORS	Direct Submission		
JOURNAL	Submitted (15-JUN-1999) Gomez-Skarmeta J.L., Biology, University of Chile, Las Palmeras 3425, Casilla 653, CHILE		
FEATURES	Location/Qualifiers		
source	1..4871 /organism="Xenopus laevis" /db_xref="taxon:8355" /dev_stage="10.5-11"		
gene	165..4715		

[illegible]

Db	549	CGTGAAGGAAATGTGGTGTGTTGTTCTCGAAGATGTGAACCATATCTCGATACGAACACG	608
OY	616	GAAGAGCTGATGAACAAAGTGTATATAGCATCTTGTGGGACCAACAGAAATTT	675
Db	609	GAACAACTGATGATACTAGTGTGTACAGCATTTTGCATGTGGGAGATATCAGCAGGATTC	668
OY	676	GTCAAAAACCTGCTGCCAAGTCTATATAGTAATAGGGGGATCTTGGTCTGGCAACCTCGC	735
Db	669	ATCAAAAACCTGCTGCCAAGGCTTTATATAGTAAGGGGG-----TGCTT	710
OY	736	AGGGGAACACCCATACCTTCATATGTGGGATCTGTGTAAACCTTTACCTATTCACAA	795
Db	711	AGCGGAATTAACCTTACCTTTAACTGTCTGTATCTGTGTAAAGCATAATGTGAATGCCAA	770
OY	796	GAGGAGGGTCTATGATTAACAGGAAGCTCATCAGAAATATGAACATATGCAATGCTTCGCT	855
Db	771	GAGGAAGGCGACGATGTGTACAGAAACATCAGAAATAGGAAGATGCAATGCAATGCTTCT	830
OY	856	GTCTCTCAACCAAGTCCATCAAGAAAGAAAGAAAGATTTGCACTGCTTCAATTTGC	915
Db	831	GTATCCCAACCAAGTCAATTAAGAAAGAAAGGTAAGATTTCCAGTCTGTTGTATGTTGT	890
OY	916	GTGGCAAGAAGATTTCCATGAAGGAAGAACCGAGTCTTCCCATCATCAGAAAGTTTACT	975
Db	891	GTGGCCAGAAAGTTCAGATGAAGAAAGAGGCCGTGTCTCTCCCATCTGAGAGTTCACA	950
OY	976	ACTGCGACGATCTCCCAAGCACAAGATCACGTCCTGTGATACAGACACATAGAGACACC	1035
Db	951	ACGGGTACGAGCTTTCAAAGTAAATATACATCTGTAGACACAACCAAGTATGAGGGCCTA	1010
OY	1036	ATGAACCAAGGCTGGAGAGACTGCTGTAGAGGTGTATTTACAGATTTCCATGCGCAGAT	1095
Db	1011	ATGAGACAGAGATGGAGAGCATGTGTGGCGCTTGTCATCCAGATTTTCATTTCCACGAT	1070
OY	1096	GAAGAGAAATGTGTCTCTATGCTATAGAGGCATCATATGAAGTACTGTGACAAAGATTG	1155
Db	1071	GATGAGAGA---ATATCTTACAGACGAAGACATACCAAGAAAGTCTCAGGAGGGGAT	1127
OY	1156	GCATTCAGTCAAAATCTATCTGTTTTTTCCTGTGATGAGCACTCTGTGTTCTGCACAAAG	1215
Db	1128	GGCAACAGTCTCTTCTTCAACCGCTTTTCTGTGTGGATGTGACACGTCTTGTGGCCAAACA	1187
OY	1216	AAGAGCAAACTCATCCGTTCTCAGACTACTAATGAACCTCACTGTATAATATCTTTACAT	1275
Db	1188	AAGAGCAGGCTCATGCGTTCCACAGACTAATATATGAACCTCCGCTGTGTATCTCTACAT	1247
OY	1276	ATGCTTCAAGAGAGCGAATGTGTGTATATATCCGATCTGACTGGACAACGATG	1335
Db	1248	GTGCTTCAGAGAGAAAGAAATGTGTTGGATTAAACAGAGCTTGGCTGGGCAACAATG	1307
OY	1336	GGGAAGCCACTGATTCCAATTAGCTGTAAAGCGCTGGCCATAGGCGCTGTGCAAGTGG	1395
Db	1308	GGAAAGACATTAATTCAGATTCACTCAGCAGATCCAGCTCATCAGCAATATGACGTGGG	1367
OY	1396	AACCCAGTCAAGACATGACCCCTCAGTAGCAATATATAATTTTCCATAAATAGGCCCAAG	1455
Db	1368	AACCCAGTCAAGATATGACCATATAGTAGCAATATATAATATAGGCCAAG	1427
OY	1456	GAACAAATGGCATGCCCCATGGCAGGTTGTGTGTTCTTGGGGAATGAACATGTGTCA	1515
Db	1428	GAACAAATGGCTTGGGAGCTGGCAGGTTGTGTGGTTCAAGGTGAATGAACCATATCTCA	1487
OY	1516	GGCATCGAAGCAACCATCCGTCAGGGGATGATATATGCACTCAAAATGAAGAGCCCTCA	1575
Db	1488	AGTTTGCAAGCAACCATCTCTCAGGGTATATATCTATGCACTTAAATGTATAGCCATCA	1547
OY	1576	CAAAAGCAGCCCTGGCATGAATTCAGAGACAGCCCACTCCATGCTTTCACACAAGGATGCG	1635
Db	1548	CATGAAGTCTCTGGCATGGGCGACGGGTCAACCTAATTTATGTGTTTCACACAAGACATGCT	1607
OY	1636	ATGAGCCCTGGATGTGCTGGACGCTCGAATCCACCAAGTCAAGTTTCCCTCGAGGA	1695

D	b	1608	GTCAAGCCAGAGAGTGGCAGAGAACTCCTCGGATTGGACACGAGTCCGTTTTCCTCCGACAGA	1607
Q	y	1696	AGCTTGCATTCCTCCCTGTGGGAGATTTCACAGACACAGAAATAGCCATAGTTATACCAAC	1755
D	b	1688	AGTTTGCATTCACCAAGTTACTGTTTGCAGTAGACAGAAATAGCCATAGTTATACCAAC	1727
Q	y	1756	AGCTTCCTCAATGACTCTTCAGGCCCCCCTCAGCGAGAGGGCAGGGGGTCTCATTTAGGTCATCG	1815
D	b	1728	AGCTCTCTTAATGACTTTCAGGCTCTTCAGAGAGGGTCAAGG---CCCTCTTGGCCCCCA	1784
Q	y	1816	TTTGCTTCCACAGACCTAAATAATGGGCAATTTGCCAAATCTCCCAAGTTAATATGAATCCT	1875
D	b	1785	TTTGCTCTCTCTGATCTGCAAAAGGTGGAAATCTTCAGACATTCGCGCTGGCAACATGAACCT	1844
Q	y	1876	CCCCACTCAGCAAGATGGGAACTTGGACTCAAAAGACTGTTTGGACTATATGGGAG	1935
D	b	1845	CCTCAGCTTAGAAAGATGGGAGACATATAGACTTAAGAAATCTTTTGGTCTGTATGAGAA	1904
Q	y	1936	CCCTCTGAAGGTACACTGTGACACAGAGAGACGCTGCCATCTGTGAGAGCAAAAGAA	1995
D	b	1905	CAGCTTGAAAGTGCACACTGTGACAGGGGGAGTCCGGCTGCCATTCAAATATGAGAGAAAGAC	1964
Q	y	1996	ACAATATGCCCCAACCTGCCCCCGGCGCTGAGACAGTGAAGAGCTGAGGGGACAGACGA	2055
D	b	1965	TGCGGTG-----AAACTGTCTAGTGTGTTTGAACAAACAGAAAGCCAAAGACGA	2015
Q	y	2056	CTGCATGACAGCAAGGGCAGACCAAACTCTGCAGCTCTGACCCACCAAAATCTGATAC	2115
D	b	2016	TTTGCTTGTATGGCAAAAGGTGACGAAAAAACTGTGAAATCTGCTCACCAACAAATCTGACC	2075
Q	y	2116	ATGAGCCCTCGCCCTTAGCCAG---CTCTTGTGGATACAAACAAAGACTCCACAGCT	2172
D	b	2076	ATGGAGGCATCAACGCTACCTAGTAATCTTTGGAGACATGAACAAAGACTCTTGCA	2135
Q	y	2173	AGCTTGC-----TGGTTGGGGTCAACATGAGAACCTCGCTCAAGGAAAGAT	2223
D	b	2136	AATTTGATCAATAATAGTATGTATGTCTGCTCAGCAATGAGAACTCACTGAAGAAAGAT	2195
Q	y	2224	AAAAATTTTGCACAGACTCTTGCAGAGACAGACTTCCCTGTGGACTTGGCCAAAGTTACA	2283
D	b	2136	AAAATTTTGCACAGACTCTTGCAGAGACAGACTTCCCTGTGGACTTGGCCAAAGTTACA	2285
Q	y	2284	GCAGAGCCACAGGCAAAAGACTGTAGCCAGAGCTCCAGCAGACAGCTTCGTGATCAGAA	2343
D	b	2286	GCAGAGCCACCGGAAAGAGCTGTAGCCAGAGAGTCCAAATAGCACAGGCCCTGTTCAGAG	2315
Q	y	2344	GTGACTATTAACAAAGACGGGTGTAGCCCCAGAAAGAAAGAAATGTACTACTTGGCTAT	2403
D	b	2316	GTCACTATAAAGCGGAACCAAGTCAGCCCCCAAAAAAAGACATGTCACTACTTGGCTATC	2375
Q	y	2404	TTTGCTAGTAAAGATGTACTTAAAGATATTTGGTTTACCAGAAATTAACCCCAACTTGA	2463
D	b	2376	TTATTAGATTAAAGATGTACTTAAAGATATTTGGTTTACCAGAAATTAACCCCAACTTGA	2422
Q	y	2464	AGACTGCAGATTAAGACAGATCTCTGCCAGTAAACAAATTAATAGCAATGAACCTGAG	2533
D	b	2483	CGAGAGCATTAACAAGGTGTAGACCTTAGACGCTGTCCAAATTTTCGGCTGTCAACACAG	2492
Q	y	2524	AAGAGAGGATGACTTTTGAAGCTGTGTGACACCACTGTGCAAGTGTGAGCTGTGACAACTTGA	2583
D	b	2493	AAGGAGGAGCCAAACTTTTGAACACACTGTGACACCACTGTGATTTGATTAATCTGGAT	2552
Q	y	2584	GAGATTTTGGATGTATTTGCAAAATATGCAATTTCCACACTTTTCCAGACAGAGGCA	2634
D	b	2553	GAAATTTTGGATGTATTTGCAAAATATGCAATTTCCACACTTTTCCAGACAGAGGCA	2612
Q	y	2644	GGCGCCCTGTGATGATTTGACACAGACCCATCATCAATGACTCTGACACTGCACACTCA	2703
D	b	2613	G-----ATGGCAACTCAGGAGGATAAACAAGCATCATGATGACTGTGACCACTGGCT	2666
Q	y	2704	GCTGAAAAAGCCCTGTGCACACTGTTTGGAGCC---AATAAAGACACTGTGGAATTTTCA	2760
D	b	2667	GGTGAACACACAGCTCTCCGCTGGGGCCCAACAGATGATGTCTCGGTATGAG	2726

QY	2761	CAGAGGCACTTTTAAATTAACCAAGACCAAGGCGAACTGGGAGGTTATTTGGCAAAACAGAT	2820
Db	2727	CAGAACATGTTTCAATTAAGCCAACTGGCCGACAGCTGGGCGCTTCCCAAAACAGAT	2786
QY	2821	TTAACCACTTGAACATTCGCAAAAGCCCAAGTGGTGTGACCTTTTCCCAATCAGA	2880
Db	2787	TTAACCACTGGAATTCATTTCCCAAAAGCCCAAGGATGACGAGTCTTTTGGCAAAATGAG	2846
QY	2881	AACGATAGTCCCTACTCAGTGAATCTCAGCCAGGAATGATGGGTAAATCAAGGATGTA	2940
Db	2847	AGCAGCGCTCCGTACACTACTGTGGCTCCAGTGTGGATTAATTAACAGGCAATGATG	2906
QY	2941	GGAACCAAGAAATTTAGGAGACAGTACACAGGAATATTTGTAAACAGTCTTCGG	3000
Db	2907	GGCAGTCAAGGCAATGTCCCTAAACAGCAGCCAGGATTTGTAGAGTGAATGGCCCTAG	2966
QY	3001	CTTACTATGCCATCTGGAGAGATGGCAGCGCAGAGTTGGCTGTGAGATCACTGTGCT	3060
Db	2967	CCACCATTAAAGCCAGAGAGCTGGGGCTCTCCAGGATCTGGCTGTGAGACAGATGGCCG	3026
QY	3061	GCTAACCCACAGTGCATGGAACCGGGCCAGTCCAAAGGAGTATATTTGGAAACCCAGCAGCC	3120
Db	3027	ACCAACAGCACTCTATGAACAGGCAATGAC-----ATGTCCCGAGTCCAACTGCTCC	3077
QY	3121	AGCATCCCCATGAGGCCCACAGCAGCCAGCTGTGGCCAAAGACAGACCTTTCAGTCTCAGGTC	3180
Db	3078	AGCATTCCTATGAGGCGCGGGCAGCCAAAGTGTGCCCTTAGGCAAGTGTCTCACTCAGCAGTA	3137
QY	3181	ATGAAATATAGGGCCATCTGAATTAAGATGAACATATGGGGGACCTCAATATAGCCAAACA	3240
Db	3138	ATGAAATATAGGGGCTTCTTGAGCTCGAATTAAGACATTAGCGGTCCCACTAACACCAGAG	3197
QY	3241	CAAGCTCCCTCAATCAGACTGCGCCATGGCCCTGGAAGCATCTGTATAGACAGAGCG	3300
Db	3198	CAGACACCCCTTAACCAACCCGACCCCTGACCTTAACCGATTTTGACTATTTAGCAACCC	3257
QY	3301	TCTTTTGCCACGCAAAACAGCAGCCATTTGGCAGTTCTCAGATGACTGTCTATGTCCA	3360
Db	3258	TCCTTCACACATCAAAACCGGCGCCATTTGGCAGTCTGCGCGAGATTTTATTTGTGTACG	3317
QY	3361	CATCTCCACACTAGATCTCCGAGATGTATGAGGAGAGCTCTCTGGACACACTGTATCTGGCC	3420
Db	3318	CCAAATGTGTGTCGAGTCTCTCTGACAGCAGATGGGAACCTTCTGTGACACACTTTATATGGCC	3377
QY	3421	TTGGCGAATTTTGATGGGCTGAGAGAGATTTGATAGACCTTAGAATACCGGAATGGTC	3480
Db	3378	TTTGGAATCTTTGACGAGCTAGAGAGATTTGACAGAGCCCTTGGAATCCAGAGATGGTT	3437
QY	3481	AGCCAGAGCCCAACAGTAAATCCAGAACGTTCTCAAGTCAGAGATTTCCAAATCATGCTG	3540
Db	3438	AGCCAGGCTCAGGCTGTGAACAGAGATCTTTGGAAAGTCTCGAATCAAACTTTATGATG	3497
QY	3541	GAGCAGAGAGCGCCCTTTTCCCAACAGCAGTATGATCTCAGGCAACAATGGCCAGGAGT	3600
Db	3498	GAGCAAAAGCGCGCTGTCTAATACCATGTGATATGCAAAATCAAGGCGCAATGGCTCAGAT	3557
QY	3601	AGCATTTCTCCCATGCAAGATCCAAACTTTCAACACCATGGGACAGCGGCTTATGTTATGCC	3660
Db	3558	AGTTTACCAACCAATGAGAGATCTCTGGTTTCAACCCTATATGGGCAACGCTCCAAAGTATGGA	3617
QY	3661	ACAATCGGTATGAGGCCAGACAGCGGGCTCAGGGCCAGCGGGCTGTAGTCAGAACAGGCA	3720
Db	3618	ATCTTGCGTATGCAAAACAGGCGCTGGGCTTACACAGCCAGACAGAAATGTGTCAAGACAGCA	3677
QY	3721	AATCAACTAAGACTTCAACTTACGATCGCTCCACAGCAGACAGAGATGGCAGCAGCACTT	3780
Db	3678	AACCAAGCTGAGATCCCAACTGACAGCACCGCTTTCAAGC---GCAAGAACCGGACAGCAATTA	3734
QY	3781	ATGAAATCAATGACAGATGTTTCCATATGTAACCTTGACATCTGAGGCTGGAGGTATAC-----	3836
Db	3735	ATGAAACCAATTAATATGTCTCAAAACATGAATCTTGACCAATATAGACCTGGAGGTATCCAGG	3794

QY	3837	-----ACACGAGGACGATTTATGATGACAGAGTGCAGCCACAGACAGAGGGAATTC	3888
Db	3795	CAGTCGAGAGAGACGACGGGCGCCATCATATGCACAGATGTTGGCACAAGGACGACAGAGATTA	3854
QY	3889	CTGAACGACGATCTTTCGACAGACAAATGCATCAGACAAACGATTCAGCAACGA	3945
Db	3855	TTGAGCCAAACGCTTACGCCAAAGCAACTTCACGACGACGACAAACAAACGA	3911
RESULT	6		
AC090280/c			
LOCUS			
DEFINITION			
ACCESSION	AC090280	86554 bp	DNA
VERSION	AC090280		
KEYWORDS	HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTLP.		
SOURCE	human.		
ORGANISM	human.		
REFERENCE			
AUTHORS	Biren, B., Linton, L., Nusbaum, C. and Lander, E.		
TITLE	Unpublished		
JOURNAL			
REFERENCE			
AUTHORS	Biren, B., Linton, L., Nusbaum, C. and Lander, E.		
TITLE	Unpublished		
JOURNAL			
REFERENCE			
AUTHORS	Biren, B., Linton, L., Nusbaum, C. and Lander, E.		
TITLE	Unpublished		
JOURNAL			
REFERENCE			
AUTHORS	Biren, B., Linton, L., Nusbaum, C. and Lander, E.		
TITLE	Unpublished		
JOURNAL			
REFERENCE			
AUTHORS	Biren, B., Linton, L., Nusbaum, C. and Lander, E.		
TITLE	Unpublished		
JOURNAL			
REFERENCE			
AUTHORS	Biren, B., Linton, L., Nusbaum, C. and Lander, E.		
TITLE	Unpublished		
JOURNAL			
REFERENCE			
AUTHORS	Biren, B., Linton, L., Nusbaum, C. and Lander, E.		
TITLE	Unpublished		
JOURNAL			
REFERENCE			
AUTHORS	Biren, B., Linton, L., Nusbaum, C. and Lander, E.		
TITLE	Unpublished		
JOURNAL			
REFERENCE			
AUTHORS	Biren, B., Linton, L., Nusbaum, C. and Lander, E.		
TITLE	Unpublished		
JOURNAL			
REFERENCE			
AUTHORS	Biren, B., Linton, L., Nusbaum, C. and Lander, E.		
TITLE	Unpublished		
JOURNAL			
REFERENCE			
AUTHORS	Biren, B., Linton, L., Nusbaum, C. and Lander, E.		
TITLE	Unpublished		
JOURNAL			
REFERENCE			
AUTHORS	Biren, B., Linton, L., Nusbaum, C. and Lander, E.		
TITLE	Unpublished		
JOURNAL			
REFERENCE			
AUTHORS	Biren, B., Linton, L., Nusbaum, C. and Lander, E.		
TITLE	Unpublished		
JOURNAL			
REFERENCE			
AUTHORS	Biren, B., Linton, L., Nusbaum, C. and Lander, E.		
TITLE	Unpublished		
JOURNAL			
REFERENCE			
AUTHORS	Biren, B., Linton, L., Nusbaum, C. and Lander, E.		
TITLE	Unpublished		
JOURNAL			
REFERENCE			
AUTHORS	Biren, B., Linton, L., Nusbaum, C. and Lander, E.		
TITLE	Unpublished		
JOURNAL			
REFERENCE			
AUTHORS	Biren, B., Linton, L., Nusbaum, C. and Lander, E.		
TITLE	Unpublished		
JOURNAL			
REFERENCE			
AUTHORS	Biren, B., Linton, L., Nusbaum, C. and Lander, E.		
TITLE	Unpublished		
JOURNAL			
REFERENCE			
AUTHORS	Biren, B., Linton, L., Nusbaum, C. and Lander, E.		
TITLE	Unpublished		
JOURNAL			
REFERENCE			
AUTHORS	Biren, B., Linton, L., Nusbaum, C. and Lander, E.		
TITLE	Unpublished		
JOURNAL			
REFERENCE			
AUTHORS	Biren, B., Linton, L., Nusbaum, C. and Lander, E.		
TITLE	Unpublished		
JOURNAL			
REFERENCE			
AUTHORS	Biren, B., Linton, L., Nusbaum, C. and Lander, E.		
TITLE	Unpublished		
JOURNAL			
REFERENCE			
AUTHORS	Biren, B., Linton, L., Nusbaum, C. and Lander, E.		
TITLE	Unpublished		
JOURNAL			
REFERENCE			
AUTHORS	Biren, B., Linton, L., Nusbaum, C. and Lander, E.		
TITLE	Unpublished		
JOURNAL			
REFERENCE			
AUTHORS	Biren, B., Linton, L., Nusbaum, C. and Lander, E.		
TITLE	Unpublished		





QY 5861 TATGTGGAATTCAGTCAAAAGAGTACTTACGATTAATAATTAGATATCTTCATAGCA 5920  
 |||||||  
 Db 45252 TATGTGGAATTCAGTCAAAAGAGTACTTACGATTAATAATTAGATATCTTCATAGCA 45193  
 |||||||  
 QY 5921 ATGTCCTCAAGAGTGTGTTTGTAAAGATATCAATGCCCTGATTAAGACCTAATTTGTA 5980  
 |||||||  
 Db 45192 ATGTCCTCAAGAGTGTGTTTGTAAAGATATCAATGCCCTGATTAAGACCTAATTTGTA 45133  
 |||||||  
 QY 5981 CTTAAGACTTTTATTTTCTTAAGACTTGTGATCTGTATTAAGTCAATTTATCAATCA 6040  
 |||||||  
 Db 45132 CTTAAGACTTTTATTTTCTTAAGACTTGTGATCTGTATTAAGTCAATTTATCAATCA 45073  
 |||||||  
 QY 6041 TATGATATGACGCGCTGTAGAGCAACATTCGTATTTTATGTTATATTCCTTCTT 6100  
 |||||||  
 Db 45072 TATGATATGACGCGCTGTAGAGCAACATTCGTATTTTATGTTATATTCCTTCTT 45013  
 |||||||  
 QY 6101 AATGACCTTGAAGAAGACTATGTTACTTAAGACGCGCACTTTTATGTTGTTT 6156  
 |||||||  
 Db 45012 AATGACCTTGAAGAAGACTATGTTACTTAAGACGCGCACTTTTATGTTGTTT 44957  
 |||||||

RESULT 7  
 AC084251/c 181403 bp DNA linear HTG 06-DEC-2001  
 LOCUS Homo sapiens chromosome 8 clone RP11-152C15 map 8, \*\*\* SEQUENCING  
 DEFINITION IN PROGRESS \*\*\* 13 unordered pieces.  
 ACCESSION AC084251 GI:17386328  
 VERSION AC084251.4  
 KEYWORDS HTG; HTGS\_PHASE1; HTGS\_FULNOP; HTGS\_ACTIVEFIN.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 181403) Nussbaum, C. and Lander, E.  
 AUTHORS Birren, B., Linton, L., Nussbaum, C. and Lander, E.  
 REFERENCE 2 (bases 1 to 181403) Human Male BAC  
 AUTHORS Birren, B., Linton, L., Nussbaum, C., Lander, E., Allen, N.,  
 Anderson, S., Barna, N., Bastien, V., Bede, F., Boguslavsky, L.,  
 Boukhalter, B., Brown, A., Burkett, G., Campopiano, A., Castle, A.,  
 Choquel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P.,  
 Dearlign, K., Dewar, K., Diaz, J., Dodge, S., Ferreira, P.,  
 Fitzhugh, M., Gage, D., Galagan, J., Gardyna, S., Goyette, M.,  
 Graham, L., Grand-Pierre, N., Hagos, B., Heath, A., Horton, L.,  
 Iller, I., Johnson, R., Jones, C., Kann, L., Karatas, A., LaRoque, K.,  
 Lamares, R., Lander, S., Lenock, J., Levine, R., Liu, C., Liu, G.,  
 Macdonald, P., Marquis, N., McCarthy, M., McEwan, P., McKernan, K.,  
 McPheters, R., Meldrum, J., Meneus, L., Mihova, T., Mlenga, V.,  
 Morrow, J., Murphy, T., Naylor, J., Norman, C. H., O'Connor, T.,  
 O'Donnell, P., O'Neill, D., Oliver, T. M., Oliver, J., Peterson, K.,  
 Pierre, N., Pisan, C., Pollara, V., Raymond, C., Rieback, M., Riley, R.,  
 Rogov, P., Rothman, D., Roy, A., Santos, R., Schuster, S., Severy, P.,  
 Sougnez, C., Spencer, B., Stange-Thomann, N., Stojanovic, N.,  
 Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,  
 Tirrell, A., Travers, M., Trigilio, J., Vassiliev, H., Viel, R., Vo, A.,  
 Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J.,  
 Zimmer, A. and Zody, M.  
 DIRECT SUBMISSION  
 Submitted (18-OCT-2000) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Dec 6, 2001 this sequence version replaced g1:14150938.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center  
 Center: Whitehead Institute/MIT Center for Genome Research  
 Center code: MIBR  
 Web site: http://www-seq.wi.mit.edu  
 Contact: sequence\_submissions@genome.wi.mit.edu  
 ----- Project Information  
 Center project name: L1326  
 Center clone name: 152\_C\_15

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 13 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 987: contig of 987 bp in length  
 1 988: gap of 100 bp  
 1 1088: contig of 17531 bp in length  
 1 1819: gap of 100 bp  
 1 18719: contig of 6222 bp in length  
 1 24941: gap of 100 bp  
 1 25041: contig of 3051 bp in length  
 1 30092: gap of 100 bp  
 1 30192: contig of 20168 bp in length  
 1 50360: gap of 100 bp  
 1 50460: contig of 25016 bp in length  
 1 75476: gap of 100 bp  
 1 7576: contig of 11371 bp in length  
 1 86947: gap of 100 bp  
 1 87047: contig of 17526 bp in length  
 1 104573: gap of 100 bp  
 1 104573: contig of 5327 bp in length  
 1 110000: gap of 100 bp  
 1 110100: contig of 15383 bp in length  
 1 125483: gap of 100 bp  
 1 125583: contig of 4186 bp in length  
 1 129769: gap of 100 bp  
 1 129869: contig of 43496 bp in length  
 1 173365: gap of 100 bp  
 1 173465: contig of 7939 bp in length.

FEATURES  
 source  
 1. 181403  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="8"  
 /map="8"  
 /clone="RP11-152C15"  
 /clone\_id="RP11-152C15" Human Male BAC  
 BASE COUNT 49970 a 40664 c 38806 g 50648 t 1315 others  
 ORIGIN

Query Match 25.9%; Score 1592.8; DB 2; Length 181403;  
 Best local Similarity 99.5%; Pred. No. 0;  
 Matches 1608; Conservative 0; Mismatches 7; Indels 1; Gaps 1;

QY 4542 ACGAATATTCGACACCTCTGAAGCACTGCTTCTGACGTACCGGCTCATTG 4601  
 |||||||  
 Db 144116 ACAAAATATTCGACACCTCTGAAGCACTGCTTCTGACGTACCGGCTCATTG 144057  
 |||||||  
 QY 4602 CTCAAAACACTTCAGCTGAGAGAGCTGCTCAATTTGTTCAACCCAGACGACTGCA 4661  
 |||||||  
 Db 144056 CTCAAAACACTTCAGCTGAGAGAGCTGCTCAATTTGTTCAACCCAGACGACTGCA 143997  
 |||||||  
 QY 4662 GCCGTTCTGCTAGAGCAGACAGGCTGGCCCTGTTCCAGGGTGGCTCAGCTG 4721  
 |||||||  
 Db 143996 GCCGTTCTGCTAGAGCAGACAGGCTGGCCCTGTTCCAGGGTGGCTCAGCTG 143937  
 |||||||  
 QY 4722 GTGCGAGAGAGAGCTGCTTCTTCTTGACAGTGAAGCTGCGATCAGACATGCTC 4781  
 |||||||  
 Db 143936 GTGCGAGAGAGAGCTGCTTCTTCTTGACAGTGAAGCTGCGATCAGACATGCTC 143877  
 |||||||  
 QY 4782 AGTCTGTTCCCTGATTCACCTTAGTGAATCTGCTCCCAAGTAATGTTG 4841  
 |||||||  
 Db 143876 AGTCTGTTCCCTGATTCACCTTAGTGAATCTGCTCCCAAGTAATGTTG 143817  
 |||||||  
 QY 4842 ACAGGCAATTTACATCCATGATGATTAATTAATGATTAATTAAGAGA 4901  
 |||||||  
 Db 143816 ACAGGCAATTTACATCCATGATGATTAATTAATGATTAATTAAGAGA 143757  
 |||||||

[illegible]

QY	5981	CTTAAGCTTTTATTTTTCGAAACCTGTAACTTCGTTAAAGTCAATTAATCAATCA	6040
Db	142676	CTTAAGCTTTTATTTTTCGAAACCTGTAACTTCGTTAAAGTCAATTAATCAATCA	142617
QY	6041	TATGATATGACGCGCGTGTAGGAACCAATTCGTGATTTATATATATATCTTCTT	6100
Db	142616	TATGATATGACGCGCGTGTAGGAACCAATTCGTGATTTATATATATATCTTCTT	142557
QY	6101	AATGAACCTTAGAAGACTACATGTACTAAGCAGGCACCTTTATATGCTGTTTTT	6156
Db	142556	AATGAACCTTAGAAGACTACATGTACTAAGCAGGCACCTTTATATGCTGTTTTT	142501

RESULT	8
AF254984	
LOCUS	71701 bp DNA linear HTG 24-OCT-2001
DEFINITION	Homo sapiens chromosome 8 clone RP4-611F18, WORKING DRAFT SEQUENCE.
ACCESSION	AF254984
VERSION	AF254984.4 GI:1635685
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE	human.
ORGANISM	Homo sapiens

**REFERENCE**

1 (bases 1 to 71701)

Wen, G., Baumgart, C., Bleichschmidt, K., Dette, M., Jahn, N., Rehder, H., Hartwig, T., Böttcher, A., Hoffmann, S., Gebhardt, M., and

Chromosome 8 genomic sequence  
unpublished  
2 (bases 1 to 71701)  
Polley, A., Wen, G., Baumgart, C., Dette, M., Jahn, N., Schlöbner, M.,  
Menzel, U., and Rosenthal, A.

JOURNAL

**COMMENT**

Sequence Quality Assessment:  
This entry has been annotated with sequence quality estimates computed by the Phrap assembly program. All manually edited bases have been reduced to quality 10. Quality levels above 40 are expected to have less than 1 error in 10,000 bp.  
Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part of this entry's ASN.1 file.

NOTE: This is a 'working draft' sequence. It currently consists of 20 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will

```

Sequencing vector: M13: 100% of reads
Chemistry: Dye-terminator Big Dye: 100% of reads
Assembly program: Phrap: version 0.990329
Consensus quality: 50971
Consensus quality: 57915 bases at least Q40
Consensus quality: 63233 bases at least Q20
Quality coverage: 3.23 x In Q20 bases; sum-of-ctrlsigs
-----
Sequence Quality Assessment:

```

Center: Institute of Molecular Biotechnology  
Center code: IMB  
Web site: <http://genome.imb-jena.de/>  
Contact: [gsctj-submit@genome.imb-jena.de](mailto:gsctj-submit@genome.imb-jena.de)  
-----  
Project Information  
Center Project name: H266  
Center Clone name: RP4-611F18

Direct Submission  
Submitted (12-APR-2000) Genome Analysis, Institute of Molecular  
Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany  
On Oct 24, 2001 this sequence version replaced gi:14327871.  
Genome Center

Men, G., Baumgart, C., Bleichschmidt, K., Dette, M., Jahn, N.,  
Lehmann, R., Menzel, U., Polley, A., Reichwald, K., Schlhabel, M.B.,  
Schudy, A., Städtiqui, R., Taudien, S., Rosenthal, A. and Platzer, M.  
Chromosome 8 genomic sequence  
unpublished  
2 (bases 1 to 71701)  
Polley, A., Men, G., Baumgart, C., Dette, M., Jahn, N., Schlhabel, M.B.,  
Schudy, A. and Rosenthal, A.

AF254984 71701 bp DNA linear HTG 24-OCT-  
Homo sapiens chromosome 8 clone RP4-611F18, WORKING DRAFT SEQUE  
20 unordered pieces.  
AF254984 AF254984.4 GI:16356859  
HTG: HTGS\_PHASE1; HTGS\_DRAFT.  
human, sapiens

TTGAACTTACGAAGACTACATGTTTACTTAAGCAGAGCAGCTTTATGTTGTTTT 6156  
 TTTT|TTTT|TTTT|TTTT|TTTT|TTTT|TTTT|TTTT|TTTT|TTTT|TTTT|  
 TTTT|TTTT|TTTT|TTTT|TTTT|TTTT|TTTT|TTTT|TTTT|TTTT|TTTT|  
 TTGAACCTTACGAAGACTACTGTTTACTTAAGCAGAGCAGCTTTATGTTGTTTT 14251  
 TTTT|TTTT|TTTT|TTTT|TTTT|TTTT|TTTT|TTTT|TTTT|TTTT|TTTT|

TTAGACTTCTTTATTTCTTAACCTGATCGTCGTATAGAATTATCATCTA 6040  
TTAAAGCCTTTATTTCTTAACCTGGATCTGCCTTAAGCATTTCTCAACGA 1426

TTGATATGACGCCCGCTGAGAACCAATCTTGATTTTATANGSTTAAATCTCTT 6100  
TTGATATGACGCCCGCTGAGAACCAATCTTGATTTTATANGSTTAAATCTCTT 7282

Page





QY 5921 ATGTCTCCTTAAGCTGTTTGTGAAGGATATCATGCTTGAATTAGACTTAATTTGAGA 5980  
 |||||||  
 DB 67558 ATGTCTCCTTAAGGCTGTTTGTGAAGGATATCATGCTTGAATTAGACTTAATTTGAGA 67617  
 |||||||  
 QY 5981 CTTAAGACCTTTTATTTTCTTAACCTTGTGATCTGCTTAAAGTCAATTATCTAATCTA 6040  
 |||||||  
 DB 67618 CTTAAGACCTTTTATTTTCTTAACCTTGTGATCTGCTTAAAGTCAATTATCTAATCTA 67677  
 |||||||  
 QY 6041 TATGATATGACGCGCTGTGAGAACCAATCTGATTTTATATGTTTATATCTTCTCT 6100  
 |||||||  
 DB 67678 TATGATATGACGCGCTGTGAGAACCAATCTGATTTTATATGTTTATATCTTCTCT 67737  
 |||||||  
 QY 6101 AATGAACCTTGAAGACCTACATGTTACTTACAGACGCGCCTTTATGCTTCTTTT 6156  
 |||||||  
 DB 67738 AATGAACCTTGAAGACCTACATGTTACTTACAGACGCGCCTTTATGCTTCTTTT 67793  
 |||||||

RESULT 9  
 LOCUS AC090280 86554 bp DNA linear HTG 23-MAY-2001  
 DEFINITION Homo sapiens chromosome 8 clone RP4-611F18 map 8, WORKING DRAFT  
 AC090280  
 AC090280  
 AC090280.2 GI:13958483  
 HTG: HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 1 (bases 1 to 86554)  
 Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
 Homo sapiens chromosome 8, clone RP4-611F18  
 Unpublished  
 2 (bases 1 to 86554)  
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,  
 Barmann,N., Bastien,V., Boguslavsky,L., Bouckhalter,B., Brown,A.,  
 Camarata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S.,  
 Collymore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,  
 Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J.,  
 Gardina,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N.,  
 Hagos,B., Heatford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,  
 Jones,C., Karata,A., LaRoque,K., Lamazares,R., Landers,T.,  
 Lechoczky,J., Levine,R., Liu,G., Maclean,C., Macdonald,P.,  
 Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,  
 McPheeters,R., Meldrum,J., Meneses,L., Mihova,T., Mlenga,V.,  
 Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,  
 O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,  
 Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,  
 Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M.,  
 Roy,A., Santos,R., Schauer,S., Schnupack,R., Seaman,S., Severy,P.,  
 Sounez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,  
 Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,  
 Travers,M., Travis,N., Trigglio,J., Vassiliou,H., Viel,R., Vo,A.,  
 Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,  
 Zembek,L., Zimmer,A. and Zody,M.  
 Direct Submission  
 Submitted (17-FEB-2001) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On May 6, 2001 this sequence version replaced gi:12957916.  
 All repeats were identified using RepeatMasker:  
 Smit,A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

TITLE  
 JOURNAL  
 COMMENT  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: W1BR  
 Web site: http://www-seq.wi.mit.edu  
 Contact: sequence\_submissions@genome.wi.mit.edu  
 Project Information  
 Center project name: L12502  
 Center clone name: 611.F.18  
 Summary Statistics  
 Sequencing vector: Plasmid; n/a; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731  
 Consensus quality: 83006 bases at least Q40  
 Consensus quality: 84782 bases at least Q30  
 Consensus quality: 85389 bases at least Q20  
 Insert size: 85000; agarose-fp  
 Insert size: 85854; sum-of-coverage  
 Quality coverage: 6.7 in Q20 bases; agarose-fp  
 Quality coverage: 6.6 in Q20 bases; sum-of-coverage  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 8 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 3967 4066: contig of 3966 bp in length  
 \* 3967 4066: gap of 100 bp  
 \* 4067 8136: contig of 4070 bp in length  
 \* 8137 8236: gap of 100 bp  
 \* 8237 12124: contig of 3888 bp in length  
 \* 12125 12224: gap of 100 bp  
 \* 12225 18749: contig of 6525 bp in length  
 \* 18750 18849: gap of 100 bp  
 \* 18850 31371: contig of 12522 bp in length  
 \* 31372 31471: gap of 100 bp  
 \* 31472 43391: contig of 11920 bp in length  
 \* 43392 43491: gap of 100 bp  
 \* 43492 71800: contig of 28309 bp in length  
 \* 71801 71900: gap of 100 bp  
 \* 71901 86554: contig of 14654 bp in length.  
 Location/Qualifiers  
 1. 86554  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="8"  
 /map="8"  
 /clone="RP4-611F18"  
 /clone\_lib="RPC1 Human PAC library 4"  
 1. 3966  
 /note="assembly-fragment"  
 4067. 8136  
 /note="assembly-fragment"  
 8237. 12124  
 /note="assembly-fragment"  
 12225. 18749  
 /note="assembly-fragment"  
 18850. 31371  
 /note="assembly-fragment"  
 31472. 43391  
 /note="assembly-fragment"  
 43492. 71800  
 /note="assembly-fragment"  
 71901. 86554  
 /note="assembly-fragment"  
 /note="assembly-fragment"  
 clone\_end: 77  
 vector\_side: right"

BASE COUNT 25686 a 17956 c 17557 g 24632 t 723 others  
 ORIGIN

Query Match 20.7%; Score 1274.2; DB 2; Length 86554;  
 Best Local Similarity 99.8%; Pred No. 0;  
 Matches 1276; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 1278 GCTTACAGAGACAGCAATGTGTGTGATGATCCGATCTGACGCAAAAGATGGG 1337  
 |||||||  
 DB 21477 GTTATACAGAGACAGCAATGTGTGTGATGATCCGATCTGACGCAAAAGATGGG 21536  
 |||||||  
 QY 1338 GAAGCCACATGATTCATTAAGCTCTTACAGACCCCTGCCATACAGCCCTTGACAGTGGAA 1397  
 |||||||  
 DB 21537 GAAGCCACATGATTCATTAAGCTCTTACAGACCCCTGCCATACAGCCCTTGACAGTGGAA 21596  
 |||||||

```

Oy 1398 CCCAGTCCGAGCATGACCTCAGTAGCAATATAATTTTCCCATAAATGGCCCAAGGA 1457
    |||||||
Db 21597 CCCAGTCCGAGCATGACCTCAGTAGCAATATAATTTTCCCATAAATGGCCCAAGGA 21656
Oy 1458 ACAATGGGCAATGCCATGGGAGTTTGTGGTTCTGGGGGAATGAACCATGTGTGAG 1517
    |||||||
Db 21657 ACAATGGGCAATGCCATGGGAGTTTGTGGTTCTGGGGGAATGAACCATGTGTGAG 21716
Oy 1518 CATGCAACCAACCACTCCTCAGAGGTAGTACATATGCATCTCAAAATGAACAGCCCTCACA 1577
    |||||||
Db 21717 CATGCAACCAACCACTCCTCAGAGGTAGTACATATGCATCTCAAAATGAACAGCCCTCACA 21776
Oy 1578 AAGAGCCCTGGCATGAATCCAGAGACAGCCCACTCCATATGCTTTACCAAGGCATCGCAT 1637
    |||||||
Db 21777 AAGAGCCCTGGCATGAATCCAGAGACAGCCCACTCCATATGCTTTACCAAGGCATCGCAT 21836
Oy 1638 GAGCCCTGGAGTGGCTGGAGCCCTCGAATCCCAACCAAGTATTTCCCTGACAGGAAG 1697
    |||||||
Db 21837 GAGCCCTGGAGTGGCTGGAGCCCTCGAATCCCAACCAAGTATTTCCCTGACAGGAAG 21896
Oy 1698 CTTGCATTTCCCTGTGGAGATTTCAGAGACAGCAAGAAATAGCCATATTTATCCACAG 1757
    |||||||
Db 21897 CTTGCATTTCCCTGTGGAGATTTCAGAGACAGCAAGAAATAGCCATATTTATCCACAG 21956
Oy 1758 CTCCCTCAATGCATCTTACAGCCCTCAGAGGAGGCGGCTCATTTAGGGTCATCGTT 1817
    |||||||
Db 21957 CTCCCTCAATGCATCTTACAGCCCTCAGAGGAGGCGGCTCATTTAGGGTCATCGTT 22016
Oy 1818 GGGCTTACACAGACCTAAAAATGGCAATTTGCAAAAATCTCCCACTTAATATGAAATCTCTC 1877
    |||||||
Db 22017 GGGCTTACACAGACCTAAAAATGGCAATTTGCAAAAATCTCCCACTTAATATGAAATCTCTC 22076
Oy 1878 CCCACTCAGCAAGATGGGAGCTTGAGACTCAAAAAGCTTTTGGACTATATGGGAGACC 1937
    |||||||
Db 22077 CCCACTCAGCAAGATGGGAGCTTGAGACTCAAAAAGCTTTTGGACTATATGGGAGACC 22136
Oy 1938 CTCTGAAGGTACAACTGGAGCAAGAGAGAGAGTGCATCCTCGAGAGCAAAAAGGAAC 1997
    |||||||
Db 22137 CTCTGAAGGTACAACTGGAGCAAGAGAGAGAGTGCATCCTCGAGAGCAAAAAGGAAC 22196
Oy 1998 AAATGACCCCAACTGCCCCCGCCGCTGAGAGATGAGAGCTGACGGGACAGAGACT 2057
    |||||||
Db 22197 AAATGACCCCAACTGCCCCCGCCGCTGAGAGATGAGAGCTGACGGGACAGAGACT 22256
Oy 2058 GCATGACAGCAAGAGGAGCAACAACTCTGAGAGTGTGACCAACCAATCTGATCAAT 2117
    |||||||
Db 22257 GCATGACAGCAAGAGGAGCAACAACTCTGAGAGTGTGACCAACCAATCTGATCAAT 22316
Oy 2118 GGAGCCCTCGCCCTTACGACCTCTTTGTGGATACAAACAAGACTCCACAGTAGCTT 2177
    |||||||
Db 22317 GGAGCCCTCGCCCTTACGACCTCTTTGTGGATACAAACAAGACTCCACAGTAGCTT 22376
Oy 2178 GCGTGGTTCTGGTCTACATGGAACCTGCTCAAGAGAGCAATAAATTTTGCACAG 2237
    |||||||
Db 22377 GCGTGGTTCTGGTCTACATGGAACCTGCTCAAGAGAGCAATAAATTTTGCACAG 22436
Oy 2238 ACTCTTGAGAGACAGAGTTCCTCGTGTGATTTGGCCAAAGTTAACAGCAAGCCACAGG 2297
    |||||||
Db 22437 ACTCTTGAGAGACAGAGTTCCTCGTGTGATTTGGCCAAAGTTAACAGCAAGCCACAGG 22496
Oy 2298 CAAAGACCTGAGGAGAGTCCAGAGCAGAGCTCTGTGATCAGAAAGTATTTAAACA 2357
    |||||||
Db 22497 CAAAGACCTGAGGAGAGTCCAGAGCAGAGCTCTGTGATCAGAAAGTATTTAAACA 22556
Oy 2358 AGAGCCGGTGAAGCCCAAGAAAGAGAGATGCACTACTGCTATTTTGTAGATTAAGA 2417
    |||||||
Db 22557 AGAGCCGGTGAAGCCCAAGAAAGAGAGATGCACTACTGCTATTTTGTAGATTAAGA 22616
Oy 2418 TGATACTTAAAGATATGTGGTTTACCAAGAAATTAACCCCAACTGTGAGAGACTGGACAGTAA 2477
    |||||||
Db 22617 TGATACTTAAAGATATGTGGTTTACCAAGAAATTAACCCCAACTGTGAGAGACTGGACAGTAA 22676

```

```

Oy 2478 GACAGATCCTGGCAGTAAACACAAAATTAATAGCAATGAAGAACTGAGAGAGAGATGAG 2537
    |||||||
Db 22677 GACAGATCCTGGCAGTAAACACAAAATTAATAGCAATGAAGAACTGAGAGAGAGATGAG 22736
Oy 2538 CTTTGAGCCTGGGTGACACAG 2556
    |||||||
Db 22737 CTTTGAGCCTGGGTGACACAG 22755

```

RESULT 10  
AF207954/c  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

Human  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE  
AUTHORS  
TITLES  
JOURNAL  
REFERENCE  
AUTHORS  
TITLES  
JOURNAL  
COMMENT

1 (bases 1 to 122924)  
Reichwald, K., Baumgart, C., Blechschmidt, K., Dette, M., Jahn, N.,  
Lehmann, R., Menzel, U., Polley, A., Schillhabel, M.B., Schudy, A.,  
Siddiqui, R., Taudien, S., Wen, G., Siebert, R., Schlegelberger, B.,  
Rosenthal, A. and Platzer, M.  
Chromosome 8 genomic sequence  
Unpublished  
2 (bases 1 to 122924)  
Reichwald, K., Wen, G., Schillhabel, M., Menzel, U., Baumgart, C.,  
Dette, M., Jahn, N. and Rosenthal, A.  
Direct Submission  
Submitted (21-NOV-1999) Genome Analysis, Institute of Molecular  
Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany  
On Jun 7, 2001 this sequence version replaced gi:8152106.  
Genome Center  
Center: Institute of Molecular Biotechnology  
Center code: IMB  
Web site: http://genome.imb-jena.de/  
Contact: gscj-submit@genome.imb-jena.de  
Project Information  
Center project name: H265  
Center clone name: RP1-92A23  
Summary Statistics  
Sequencing vector: M13; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.990329  
Consensus quality: 112126 bases at least Q40  
Consensus quality: 116396 bases at least Q30  
Consensus quality: 119383 bases at least Q20  
Quality coverage: 5.23 x in Q20 bases; sum-of-contigs

Sequence Quality Assessment:  
This entry has been annotated with sequence quality estimates computed by the Phrap assembly program. All manually edited bases have been reduced to quality 10. Quality levels above 40 are expected to have less than 1 error in 10,000 bp. Base-by-base quality values are not generally visible from the Genbank flat file format but are available as part of this entry's ASN.1 file.

\*\*\*\*\*  
NOTE: This is a 'working draft' sequence. It currently consists of 8 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 8065: contig of 8065 bp in length  
\* 8066 8165: gap of unknown length  
\* 8166 46558: contig of 38393 bp in length





```
repeat_region /rpt_family="AluSx"
                complement(21513..21758)
repeat_region /rpt_family="MIR"
                21944..22017
repeat_region /rpt_family="AT-rich"
                complement(22060..22186)
repeat_region /rpt_family="MIR"
                complement(22543..22843)
repeat_region /rpt_family="AluSc"
                complement(22941..23128)
repeat_region /rpt_family="MER3"
                complement(23688..23729)
repeat_region /rpt_family="L1PA7"
                complement(24557..24637)
repeat_region /rpt_family="MIR3"
                complement(25039..25344)
repeat_region /rpt_family="AluSx"
                26351..26490
repeat_region /rpt_family="L1MA2"
                26615..26648
repeat_region /rpt_family="(CAT)n"
                complement(27973..28134)
repeat_region /rpt_family="MIR"
                complement(28392..28983)
repeat_region /rpt_family="AluJo"
                32586..32622
repeat_region /rpt_family="AT-rich"
                32738..32805
repeat_region /rpt_family="MIR"
                36137..36181
repeat_region /rpt_family="GA-rich"
                complement(36667..36800)
repeat_region /rpt_family="AluJo"
                complement(37421..37594)
repeat_region
```

Query Match 20.7% Score 1274.2: DB 9: Length 124630:  
Best Local Similarity 99.8% Pred. No. 0:  
Matches 1276; Conservative 0; Mismatches 3; Indels 0; Gaps 0:

```
QY 1278 GCTTCACAGAGAGCAGAAATGTGTGTGATGATCCGATCTGACTGACAAACGATGGG 1337
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 13973 GTTATACAGAGACAGAAATGTGTGTGATGATCCGATCTGACTGACAAACGATGGG 13914

QY 1338 GAAGCCACTGAAATTCATTAAGCTCTTAACAGCCTGCCATCAGGCCCTGTGACATGGAA 1397
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 13913 GAAGCCACTGAAATTCATTAAGCTCTTAACAGCCTGCCATCAGGCCCTGTGACATGGAA 13854

QY 1398 CCCAGGTCAGAGATGACCTCCAGTGAATATTAATTTTCCATTAATGGCCCAAGGA 1457
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 13853 CCCAGGTCAGAGATGACCTCCAGTGAATATTAATTTTCCATTAATGGCCCAAGGA 13794

QY 1458 ACAATATGGGATGCCATGGGAGGTTGGTGTCTGGGGAAATGAACATGTGCAG 1517
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 13793 ACAATATGGGATGCCATGGGAGGTTGGTGTCTGGGGAAATGAACATGTGCAG 13734

QY 1518 CATGCAAGACACCATCTCTCAGGGTATGTAATCTATGCACTCAAAATGAACGCCCTCACA 1577
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 13733 CATGCAAGACACCATCTCTCAGGGTATGTAATCTATGCACTCAAAATGAACGCCCTCACA 13674

QY 1578 AAGCACCCTGGCATGAATTCAGAGAGCCACCTCCATGCTTCAACCAAGGATGGCAT 1637
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 13673 AAGCACCCTGGCATGAATTCAGAGAGCCACCTCCATGCTTCAACCAAGGATGGCAT 13614

QY 1638 GAGCCCTGAGTGTGAGAGCCCTCGAATCCACCCAGTCAATTTTCCCTCAGAGAG 1697
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 13613 GAGCCCTGAGTGTGAGAGCCCTCGAATCCACCCAGTCAATTTTCCCTCAGAGAG 13554

QY 1698 CTTCGATTCCTCCTGTGGAGATTGGACAGACACAGAAATAGCCATAGTTATACCAACAG 1757
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 13553 CTTCGATTCCTCCTGTGGAGATTGGACAGACACAGAAATAGCCATAGTTATACCAACAG 13494

QY 1758 CTCCTCAATGCACTTCAGGCCCTCAGCAGAGGGGCAAGGGGTCTATTAGGTCATCGTT 1817
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

```
DB 13493 CTCCTCAATGCACTTCAGGCCCTCAGCAGAGGGGCAAGGGGTCTATTAGGTCATCGTT 13434

QY 1818 GGGTTACACAGACCTAAATAATGGGCAATTTGGCAAACTCCCGAGTTAAATGAAATCTCC 1877
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 13433 GGGTTACACAGACCTAAATAATGGGCAATTTGGCAAACTCCCGAGTTAAATGAAATCTCC 13374

QY 1878 CCCACTCAGCAGATGGGAAGCTTGGACTCAAAAGACTGTTTGGACTATATGGGAGCC 1937
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 13373 CCCACTCAGCAGATGGGAAGCTTGGACTCAAAAGACTGTTTGGACTATATGGGAGCC 13314

QY 1938 CTCCTGAGGTACAACTGGAGCAACAGAGAGCAGCTCCGATCCGAGGAGCAAAAGGAAC 1997
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 13313 CTCCTGAGGTACAACTGGAGCAACAGAGAGCAGCTCCGATCCGAGGAGCAAAAGGAAC 13254

QY 1998 AATGACCCCAACCTGCCCCCGCGTGAAGCAGTGAAGAGCTGAGGGCAGAGCAGACT 2057
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 13253 AATGACCCCAACCTGCCCCCGCGTGAAGCAGTGAAGAGCTGAGGGCAGAGCAGACT 13194

QY 2058 GCATGACAGCAAAAGGCGAGACCAAACTCCTGACGCTGTGACCAACCAATCTGATCAGAT 2117
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 13193 GCATGACAGCAAAAGGCGAGACCAAACTCCTGACGCTGTGACCAACCAATCTGATCAGAT 13134

QY 2118 GGAAGCCCTGGCCCTTAGCCAGCTCTTGTGCGATTAACAAACAAAGACTCCACAGTACT 2177
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 13133 GGAAGCCCTGGCCCTTAGCCAGCTCTTGTGCGATTAACAAACAAAGACTCCACAGTACT 13074

QY 2178 GCGTGGTTCTGGGTACACATGGAACCTGCGTCAAGAGAAAGCATAAATTTTGCACAG 2237
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 13073 GCGTGGTTCTGGGTACACATGGAACCTGCGTCAAGAGAAAGCATAAATTTTGCACAG 13014

QY 2238 ACTCTTGACAGACAGCATTTCCCTGTGACCTTGGCACTTGGCCAAAGTTAACAGCAAGCCACAG 2297
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 13013 ACTCTTGACAGACAGCATTTCCCTGTGACCTTGGCACTTGGCCAAAGTTAACAGCAAGCCACAG 12954

QY 2298 CAAGAAGCTGAGGCGAGAGTCCAGACAGCAGCTCCTGATCAAGAGTACTATTATTAACA 2357
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 12953 CAAGAAGCTGAGGCGAGAGTCCAGACAGCAGCTCCTGATCAAGAGTACTATTATTAACA 12894

QY 2358 AGAGCCGCTGAGGCCCAAGAGAAAGAAATGACACTCTTCCGATATTTCTGATTAAGA 2417
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 12893 AGAGCCGCTGAGGCCCAAGAGAAAGAAATGACACTCTTCCGATATTTCTGATTAAGA 12834

QY 2418 TGATTAATAAGATATTTGTTTACCAAAATTAACCCCAACTTGAAGACTGACAGTAA 2477
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 12833 TGATTAATAAGATATTTGTTTACCAAAATTAACCCCAACTTGAAGACTGACAGTAA 12774

QY 2478 GACAGATCCTGCCATTAACAAATAATTAATAGCAATGAAGAACTGAGAGAGAGATGAG 2537
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 12773 GACAGATCCTGCCATTAACAAATAATTAATAGCAATGAAGAACTGAGAGAGAGATGAG 12714

QY 2538 CTTTGAGCCTGTGACAG 2556
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 12713 CTTTGAGCCTGTGACAG 12695
```

```
RESULT 12
AF323989 4958 bp mRNA linear VRT 21-FEB-2001
LOCUS Dario reio transcriptional intermediary factor 2 mRNA, complete
DEFINITION cds
ACCESSION AF323989
VERSION AF323989.1 GI:13022009
KEYWORDS
SOURCE
ORGANISM
Dario reio
zebrafish.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
REFERENCE
1 (bases 1 to 4958)
Tan, J.H., Kara, A.U. and Chan, W.K.
Molecular cloning and characterization of zebrafish nuclear
receptor coactivator TIF2
JOURNAL
Unpublished
```

REFERENCE 2 (bases 1 to 4958)  
 AUTHORS Tan, J. H., Kara, A. U. and Chan, W. K.  
 TITLE Direct Submission  
 JOURNAL Submitted (25-NOV-2000) Biological Sciences, National University of  
 Singapore, Lower Kent Ridge Road, Singapore 119260, Singapore  
 FEATURES  
 source Location/Qualifiers  
 1. 4958  
 /organism="Danio rerio"  
 /db\_xref="taxon:7955"  
 196..4713  
 /note="nuclear receptor coactivator; TIF2/GRIPI"  
 /codon\_start=1  
 /product="transcriptional intermediary factor 2"  
 /protein\_id="AAK11608.1"  
 /db\_xref="GI:13022010"

## CDS

/translation="MSAVGERSSDPARSEAKREKREPSDLGSPKSTKRNHEHS  
 KYLEBELLEIPANNDNDNFVAKPDKCALIKETVKQIRKIDKEDKAAANDEYOKAD  
 VSTGOSVIDKDALGPMMLBALDFFVNMENIVSENVTOYLRYNOBELMNTSY  
 YSILHVDHAEFIKNLPSKSHVNGVPMWSENPRNSHTFNCRLMVLNHSAEETDHE  
 AOOKYETMOCFAVSEPKSKEEGEDFOSCLICVRRVPMKERPMLPTQESITTRDQD  
 GKITSIDTSLIRASMKRMEDIVRRCIORFHLNDQDISFAKHOCVIRHGAESPFI  
 XRESIDGTIVSAHTKSKLYRSSSTNEPOLYMSHLIIQRBPVPCGMAISDGNATMRK  
 PMPMSPNTAGSCTPQGDATISSNTSTPSPGAKKEFGAMHRECCPGTMSATM  
 QAAIPQSGYPLKLSFGSPGMLSPRHASPDVAGSPRLPPQFSPASLHSPAM  
 CSTSGNGGAGANHYGTSSSLNALQALSCBHSQSLGSPDKLGSPPAHSVAVS  
 HHLINKSVSEFGEONOSRGTPGODEGIDLPDEKNGIGOFENLGSPOSLRDN  
 KSHTKLIQLLTTRTEPIESTSPPMAGCEPCCKDGKNGMGSGGNGSHATSLRK  
 HKLIHLRLONSSSPVDLAKLTAETKELCLODSAGTAPBELAKOEPVSPKRDNA  
 LKRLYLDKDNVLKKGIKMEPEIKKEEGKMTGIKETKDGGIDREPESELDIIDL  
 LIDNOSQGLTDSKPLSPSAVDKQSIINLIQWTEGSGANMSPQOORALQOTAVSQN  
 PPARPCQGPRAPIRVRSISLDMNP PKGAGQGYPMNNNPYSIMOQGMGNHNAV  
 NOPMNAVAGVNGNPRVGMQODMGAGTGVSAASAPATAPAGOSHLPQGLHSRV  
 PNPATGPMRPNSGRPMLOPOMMANOSNMMDGALLSOLTSALKDQGEIIR  
 VMIEPIYPRGNOSSTKSGSOEYLCPPASBEGPDEBAILSOLTSALKDQGEIIR  
 ALGIPALVGAQPEVDQPEQDPSMMMDKPPRYGQOYASSPANNQOAGYTPMDA  
 TFHGLQOMQORRGYPMLRMQARPLRGGVVPQNPNTLRLOJLHRLQAOONRPMAN  
 OMGVSNMNLPLRPVNPQGTINAKMLAQRRLSHLRLQJLDOORQOQOOR  
 SLAMRAGLTMPMMAAAGMGANPNRPIPOANLOAFYPRYGTGSGSPSTSP  
 FSPGPMPLPAPOLISHSAMGSLANQCMGLSLSQCFAYMPOMGHAFOPSSG  
 MSQSDGREGGATPOSLMSPRKHGHOSTMPOMOAGCNSSFOSSPDINGMPQGININ  
 SMFTQDSPPQFSDOANNNTNGNMLNVMAANSSNMGMGNOMSMSSMNSGSGM  
 ANMGPEQKCYC"  
 247..453  
 /note="Region: basic helix-loop-helix DNA-binding domain"  
 568..747  
 /note="Region: PAS A domain"  
 1000..1185  
 /note="Region: PAS B domain"  
 2083..2097  
 /note="Region: NR Box"  
 2251..2265  
 /note="Region: NR Box"  
 2419..2433  
 /note="Region: NR Box"  
 4123..4139  
 /note="Region: HAT-like domain"  
 BASE COUNT 1284 a 1405 c 1343 g 926 t  
 ORIGIN

Query Match 20.1%; Score 1238.4; DB 5; Length 4958;  
 Best Local Similarity 60.8%; Pred. No. 0;  
 Matches 2515; Conservative 0; Mismatches 1411; Indels 210; Gaps 22;

QY 146 GCTGATATGTGTTCAAGATGATGGATGAGAGAAATACCTGACCCCTCCAGGCGAG 205  
 Db 179 GGTGATATGATGTGATGATGCTGTCGGGAAACCTGCTGACCCCTCCAGGCTCAG 238  
 QY 206 AGACAAGAAAGCGCAAGGATGTC---CTGACCAACTTGGACCCAGCCCAAGGAACA 262  
 Db 239 AGGCTGAGAAAGCGCAAGGATGTCAGACCTCTGGGCTCAAGTCCCAAGGAAGTA 298  
 QY 263 CTGAAAAACGTAAATCTGAACAGAAAAATAATATATAGAACTTGCAAGTTGATT 322

Db 299 CAGAAAAACGTAAATCTGAACAGAAAAATAATATATAGAACTTGCAAGTTGATT 358  
 QY 323 TTGCAAAATTTAATGATATATAGCAAACTTAACTTCAAACTGACAAATGTGCATTTAA 382  
 Db 359 TCGCAAACTTCAATGATATATAGCAAACTTCAAACTGACAAATGTGCATTTAA 418  
 QY 383 AAGAACTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 442  
 Db 419 AGGAAACCGTAAACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 478  
 QY 443 TAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 502  
 Db 479 AGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 538  
 QY 503 CGCTGGGCGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 562  
 Db 539 CCTTGGACCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 598  
 QY 563 GCAACGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 622  
 Db 599 GCACATTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 658  
 QY 623 TGATGACAAAGTGTATATATAGCATCTTGCATGTTGGGACACACGGAATTTGTCAAA 682  
 Db 659 TGATGACAAAGTGTATATAGCATCTTGCATGTTGGGACACACGGAATTTGTCAAA 718  
 QY 683 ACTGCTGCCAAAGTCTATATAGCATCTTGCATGTTGGGACACACGGAATTTGTCAAA 742  
 Db 719 ACTCTCTCTTAAGCACACGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 778  
 QY 743 ACAGCATACCTTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 802  
 Db 779 ACAGCATACCTTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 837  
 QY 803 GTCAATGATACCAAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 862  
 Db 838 -----CAGATCTCAGAGGACGACAGCAAGAAATATGACATGATGATGATGATGATGAT 892  
 QY 863 AACCAAGTCCATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 922  
 Db 893 AACCAAGTCCATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 952  
 QY 923 GAAGAGTCCCATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 982  
 Db 953 GAGAGTCCCATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1012  
 QY 983 AGATCTCCAGGCAAGATCAAGATCAAGATCAAGATCAAGATCAAGATCAAGATCAAGATCAAG 1042  
 Db 1013 AGGACTTCCAGGCAAGATCAAGATCAAGATCAAGATCAAGATCAAGATCAAGATCAAGATCAAG 1072  
 QY 1043 CAGGCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1102  
 Db 1073 CAGGCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1132  
 QY 1103 AATCTGTCTCTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1162  
 Db 1133 A--CATTTCTTCCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1189  
 QY 1163 GTCAAACTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1222  
 Db 1190 GTCCATCTACCGGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1249  
 QY 1223 AACTATCCGTTCTCAGATCAAGATCAAGATCAAGATCAAGATCAAGATCAAGATCAAGATCAAG 1282  
 Db 1250 AACTATCCGTTCTCAGATCAAGATCAAGATCAAGATCAAGATCAAGATCAAGATCAAGATCAAG 1309  
 QY 1283 ACAGAGACCAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1339  
 Db 1310 AGAGGAGACCAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1369  
 QY 1340 AGCCATGATATCAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1399



Db 1370 AACCCATGAACCCCATGAGCTCTCCAAACACAGC-----GGGTTCCCTGTCAGCTCCCC 1423  
OY 1400 CAGGTGAGACATGACCCCTCAGTAGCAATTAATTTTCCCATTAATGAGCCAAAGAAC 1459  
Db 1424 AAGGGAGAGAGCCACCATCAGACACACATCCACTTTCCCTCCCGGTGCCCA 1483  
OY 1460 AATGGGATGCCCATGGGAGTTTGTGTGTTCTGGGGGATGAACCATGTCTAGGCA 1519  
Db 1484 AGGAGGCTGGTGACATGCACCCGTTGGCGCCGGGAACCATGAGCCATTCTGGCAACCA 1543  
OY 1520 TGCACCAACCCACTCTCCAGGGTAGTAATGCAATGCACCAAAATGACAGCCCTCCACAA 1579  
Db 1544 TGTAAACGGCCACCTCTCAGGAGCGGGGTATCCTGTAAGCTGACAGCCGCTCAG 1603  
OY 1580 GCAGCCCTGGCATGAATCCAGACAGCCACCTGCATGCTTTCACCAAGGATCGCATGA 1639  
Db 1604 GTAGTCCGGG-----CATGCTTCTCCACGTCACACAGGCCA 1639  
OY 1640 GCCCTGAGTGGCTGAGCCCTCGAATCCACCCAGTCAGTTTCCCTGCAGAGCT 1699  
Db 1640 GTCTGAGTAGAGCATCTCCACGCTTACCCCGCCACAGTTCTCCAGCGGAAAGCC 1699  
OY 1700 TGCATTCCTCTGGGAGTTTTCAGACAGA-----CAGGAATTA 1738  
Db 1700 TCCATTACCTGCGACATGTGTGACTAGCACTAGTGTGCAATGGGGGAGCGAGGCCCA 1759  
OY 1739 GCCATAGTTATACCAACAGCTCCCTCAATGACACTTGCAGCCCTCAGCAGAGGCGACGGG 1798  
Db 1760 ACCATGTGTACACACAGCATCTTCAATGCTTTTCAAGCGGCTTACGCAATGCGCATGGCG 1819  
OY 1799 TCTCATTTAGGTCATGTGTGGCTTCAACAGACCTAAAAATGGSCAATTTGCAAAACTCC 1858  
Db 1820 TAAGCCACGAGCAATCTCTGGGATCCGCTGATGAAAGTTAGGTGCCCGCCAGTCGCAATT 1879  
OY 1859 CAGTTAATATGATTCCTCCCTCAGCAGATGGGAACTTGGACTCAAAAGACTGTT 1918  
Db 1880 CGGTACACAGTAACTCATCCTAATTAATTAATAATGAGCCGTACTGTAGTCA-----T 1930  
OY 1919 TTGACATATATGGGAGCCCTCTGAAGGTACAACTGAGACAGCAGAGAGAGCTGCATC 1978  
Db 1931 TCGGAGAGCGAGAACCAATGAGACAGAGGACCCCTGACAGATGAAGGATTTGATGTGC 1990  
OY 1979 CTGGAAGAGCAAAAGAAACAATGACCCCACTGCCCGCGCGGTGAGCATGTGAGAGAG 2038  
Db 1991 CAAGAATGAAAAAG-----GCAACTCGGGCGAGTTGCGCAACCTGAGCGCA 2038  
OY 2039 CTGACGGGAGAGCAGACATGATGACAGACAGAGGCAACCAACTCTCAGCTGCTGA 2098  
Db 2039 GTATCTCTCAAGCCGACCTGCGGACACACAAAGTCAACACTTACCTCTCAGCTCTCA 2098  
OY 2099 CCACCAATCTGATCAGATGGAG-----CCCTCGCCCTTAGCCAGCTCTTTGTGCG 2149  
Db 2099 CCACCTAAATGAGCCTATGATGACCTTACCTTCAACCCCTCAATGGCGGTGGAAACCTG 2158  
OY 2150 ATACAAACAAAGACTCCACAGGTA-----GCTTGCCTGGTTCTGGGCTTACAC 2197  
Db 2159 GGTGTAAAGCGGCGGTGCGAGGAATGTGCGATGGGTTCTCAAGGGGGAACGGGCTAC 2218  
OY 2198 ATGAAGACCTGCTCAAGGAGAGATAAATTTTTCAGACACTTTCAGAGCAGCAGTT 2257  
Db 2219 ATGCTACTCTCTGTAAGGAGAAACACAAATCTCTTACAGACTCTCTCAGAAAGCTCGT 2278  
OY 2258 CCCCTGTGACTTGGCCAACTTAACAGCAGACAGCAGGCAAGGCAAGCTTGGCAGAGT 2317  
Db 2279 CGGCTGTGATTTGGCCAAACTTACAGCTGAGGCAAGGGAAGAGCTGTGTCAGATTT 2338  
OY 2318 ---CCAGCAGCAGACTCTGGATCAGAGTGAATTAACAAGAGCCGCTGAGCCCA 2374  
Db 2339 CTGCTGCGGAGCGGCGGTGCGAGTTGGCCTTAACAAGAGCCGCTGACAGTCCCA 2398  
OY 2375 AGAAGAAAGATGACACTCTTGGCTATTTGGTAGTAAGATGATACTAAAGATATTG 2434  
Db 2399 AGAAAAAGCAATGCTTGTGCTGCTACTGATGATTAAGATGACAAAT----- 2448

---

OY 2435 GTTTACCAAAATTAACCCCAACTTGGAGACTGGACAGTAGACAGATCTGCCAGTA 2494  
Db 2449 -----GTACTCAAGGTTAAAGTATTTAAATAGAGCCGTGGAGAGA 2488  
OY 2495 ACACAAATTAATATGCAATGAAGAACTGACAGAGAGATGAGCTTTGAGCCTGTGAGCC 2554  
Db 2489 TCAGAGAGAGAGGGGGCAAGATGCTGGGATTAAGCCAGAAATAATGATGTGGCTAGG 2548  
OY 2555 AGCCTGGCAGTGAAGCTGGACAACTTGGAGAGATTTTGGATGATTTTGCAGAAATGCAAT 2614  
Db 2549 ACCGTGTGAGAGCCGTCAAGTGAAGTGAATGATATCTCTGATGATCTTCCAGAAAGTCA 2606  
OY 2615 TACCAAGCTTTTCCAGACAGAGGCGAGGCGCCCTCTGTGATCAGTTGACAAAGCAAG 2674  
Db 2607 -GCCGGCCGTTCACCTGACTGACTCCAGGCTGTGTCTGTGCTTGTGCTGGACAAACAGT 2665  
OY 2675 CCATCATCAATGACTCTGATGCAACTCAGAGCTGAAACAGCCCTGTCAACCTGTTGGAG 2734  
Db 2666 CAATCATCAATGACATCTGACAGATGACAGAGAGAGCGGAGCTTAACATGATGCTCAAC 2725  
OY 2735 CCCGAAACAGCAGCTG---CGAATTTACAGAGCACTTTTAATAACCCACAGCAGGGC 2791  
Db 2726 AGCAGAGAGCATGCGACTGCACTGACGTACGCCACAGAAATTTCCCTGCTCCAGCGCCGCT 2785  
OY 2792 AACTGGCAGGTTATTTGCCAAACCAAGAAATTTACACTTGACATCAGATTGCAAAAGCCAA 2851  
Db 2786 AGCCTGCGAGAGCTCTCCGGGTGAGAAAGCGTGTCTGTGACATGAAATGCTCTTAAG 2845  
OY 2852 CTGTGCTGAGCTTTCCCAACCAATCAGAAACAGTGTGCTTACTGTAGTATCTCAGC 2911  
Db 2846 GCGCTCAGGCTGATATCTTATGATGAGAAACAAACCCCTCACTCAATATGACACAC 2905  
OY 2912 CAGGAATGATGGTAAATCAAGGATGATAGAAACCAAGAAATTTAGGAAACAGTAGAGA 2971  
Db 2906 AAGGCATGATGGAAACCAACGCCGTATGCCCCATATAGCCCAATGATGTGAAG----- 2959  
OY 2972 CAGGAATGATGGTAAACAGTCTTCTCGGCTACTATGCCATCTGAGAGATGGAGCCAGCC 3031  
Db 2960 CAGGAGTCAATGGGTAAATATGATGCTCCCGGTAGGACATGACAGAGAGGCTGGGGTGTCTC 3019  
OY 3032 AGAGTTGCGCTGTGAGAGTCACTGTGTCTACCAACAGTGCATGAACCGG----- 3085  
Db 3020 AGGCTACCGTGGCAGCGCAGCTTCACTGCACTGCTGCTCCGCTGCGCAGCAGCTCTC 3079  
OY 3086 -----CAGTCCAAAGAGTATGATTTGGAACCCAGCAGCAGCAGCATCCCATGA 3133  
Db 3080 TCCCGAGAGGGGACATGCGACAGCGCATGTGTGCCAAACCTGTGCCAAGAAATGCCATGA 3139  
OY 3134 GGGCCAGCAGCAGCTGTGGCCAAAGACAGAGCTTCAAGTCAAGTCAATGATATATAGGGC 3193  
Db 3140 GACCCAAACAGCCAGCCTGAGCCAGACCAATGCTGCAACCGCAGATGAGGCCCAATGCTC 3199  
OY 3194 CATCTAATTAGATGAACATGGGGGACCTGATATAGCCAAACAAGACCTCTCCAA 3253  
Db 3200 AGTCAAAATGAGCAATGGGATGGATGGCTGCGACCAAGTTTCCCGACAAACAGCCGCCA 3259  
OY 3254 ATCAGACTGCCCATGAGCCCTGAAGCATCTGCTTATAGCAAGGCGGTCTTTTGGCCAGCC 3313  
Db 3260 ATCAGAGCGCTCATGAGCCAGACAGTGAATGATGCTTATAGAGCGGCGGTGGGAAC 3319  
OY 3314 AAAACAGGAGCCATTTGGCAGTTCTCAGATACACTTGTATGCCATCTGCAAGCTG 3373  
Db 3320 AGAGCAGTTTCAATGATGAGGCGGCTCAGAGAGATGTGTGTGCC---TCTGCGCTCTG 3376  
OY 3374 AGTCTCGAGTGTGAGGAGCTCTCTGAGCAACAGTGTATCTGCGCTTGGGCAATTTTG 3433  
Db 3377 AAGGTCGGCTGTATGAGGGGGCTTTACTAGCCAGCTTACTCTGAGCCCTTCAAAATTTTG 3436  
OY 3434 ATGCGCTGAGAGATTTGATAGAGCTTATAGAAATTAACCAAGTGTGACAGCAGCAAG 3493  
Db 3437 ATGCTCTGAGAGATTTGACAGCGTGTGCGCTTGTGCAATTCAGCGCTATGTGGAGCAGGCTCAGC 3496



```

* 247465 256788: contig of 9324 bp in length
* 256789 256808: gap of unknown length
* 256809 265588: contig of 9750 bp in length
* 265589 266578: gap of unknown length
* 266579 275232: contig of 8654 bp in length
* 275233 275252: gap of unknown length
* 275253 281450: contig of 6198 bp in length
* 281451 281470: gap of unknown length
* 281471 289954: contig of 8484 bp in length
* 289955 289974: gap of unknown length
* 289975 294111: contig of 4137 bp in length
* 294112 294131: gap of unknown length
* 294132 297704: contig of 3573 bp in length
* 297705 297724: gap of unknown length
* 297725 301771: contig of 4047 bp in length
* 301772 306666: contig of 4875 bp in length
* 306667 306686: gap of unknown length
* 306687 312494: contig of 5808 bp in length
* 312495 312514: gap of unknown length
* 312515 316460: contig of 3946 bp in length
* 316461 316480: gap of unknown length
* 316481 321689: contig of 5209 bp in length
* 321690 321709: gap of unknown length
* 321710 322020: contig of 311 bp in length
* 322021 322040: gap of unknown length
* 322041 325449: contig of 3409 bp in length
* 325450 325469: gap of unknown length
* 325470 329359: contig of 3890 bp in length
* 329360 329379: gap of unknown length
* 329380 331940: contig of 2561 bp in length
* 331941 331960: gap of unknown length
* 331961 337339: contig of 5379 bp in length
* 337340 337359: gap of unknown length
* 337360 340919: contig of 3560 bp in length
* 340920 340939: gap of unknown length
* 340940 343973: contig of 3034 bp in length
* 343974 343993: gap of unknown length
* 343994 347937: contig of 3944 bp in length
* 347938 347957: gap of unknown length
* 347958 350294: contig of 2337 bp in length
* 350295 350314: gap of unknown length
* 350315 354429: contig of 4115 bp in length
* 354430 354449: gap of unknown length
* 354450 357126: contig of 2677 bp in length
* 357127 357146: gap of unknown length
* 357147 359807: contig of 2661 bp in length
* 359808 359827: gap of unknown length
* 359828 361451: contig of 1624 bp in length
* 361452 361471: gap of unknown length
* 361472 364041: contig of 2570 bp in length
* 364042 364061: gap of unknown length
* 364062 367340: contig of 3279 bp in length
* 367341 367360: gap of unknown length
* 367361 370079: contig of 2719 bp in length
* 370080 370099: gap of unknown length
* 370100 372710: contig of 2611 bp in length
* 372711 372730: gap of unknown length
* 372731 375148: contig of 2418 bp in length
* 375149 375168: gap of unknown length
* 375169 376342: contig of 1374 bp in length
* 376343 376362: gap of unknown length
* 376363 376568: contig of 1096 bp in length
* 376569 377678: gap of unknown length
* 377679 379356: contig of 1678 bp in length
* 379357 379376: gap of unknown length
* 379377 380421: contig of 1045 bp in length
* 380422 380441: gap of unknown length
* 380442 381556: contig of 1115 bp in length
* 381557 381576: gap of unknown length
* 381577 382472: contig of 896 bp in length
* 382473 382492: gap of unknown length
* 382493 384228: contig of 1736 bp in length

```

## FEATURES

source

```

1..404593
/organism="Mus musculus"
/strain="C57BL6/J"

```

Query Match 17.4%; Score 1071.2; DB 2; Length 110000;  
 Best Local Similarity 91.0%; Pred. No.

```

* 384229 384248: gap of unknown length
* 384249 385464: contig of 1236 bp in length
* 385465 385504: gap of unknown length
* 385505 386750: contig of 1246 bp in length
* 386751 386770: gap of unknown length
* 386771 388546: contig of 1776 bp in length
* 388547 388566: gap of unknown length
* 388567 389709: contig of 1143 bp in length
* 389710 389729: gap of unknown length
* 389730 391387: contig of 1658 bp in length
* 391388 391407: gap of unknown length
* 391408 393781: contig of 2374 bp in length
* 393782 393801: gap of unknown length
* 393802 395091: contig of 1290 bp in length
* 395092 395111: gap of unknown length
* 395112 396821: contig of 1710 bp in length
* 396822 396841: gap of unknown length
* 396842 398110: contig of 1269 bp in length
* 398111 398130: gap of unknown length
* 398131 400076: contig of 1946 bp in length
* 400077 400096: gap of unknown length
* 400097 401359: contig of 1263 bp in length
* 401360 401379: gap of unknown length
* 401380 401674: contig of 295 bp in length
* 401675 401694: gap of unknown length
* 401695 403707: contig of 2013 bp in length
* 403708 404593: contig of 866 bp in length.
403728
Location/Qualifiers

```

Result No.	Score	Query Match	length	DB	ID	Description
1	4248	55.7	1505	13	Q98TW1	Q98TW1 brachydanio
2	2736	35.9	1420	4	Q9UPG4	Q9UPG4 homo sapien
3	2734.5	35.8	1417	4	Q9UPG7	Q9UPG7 homo sapien
4	2729	35.8	1424	4	Q9Y6Q9	Q9Y6Q9 homo sapien
5	2711	35.5	1412	4	Q9UPC9	Q9UPC9 homo sapien
6	2698.5	35.4	1391	13	Q575J9	Q575J9 xenopus laevis
7	2511	32.9	1398	11	Q09000	Q09000 mus musculus
8	2175	28.5	1441	4	Q43792	Q43792 homo sapien
9	2174	28.5	1399	4	Q43793	Q43793 homo sapien
10	2159	28.3	1441	4	Q00150	Q00150 homo sapien
11	2137	28.0	1405	11	P70366	P70366 mus musculus
12	2108	27.6	1405	11	Q61202	Q61202 mus musculus
13	2106	27.6	1447	11	P70365	P70365 mus musculus
14	1684	22.1	1082	11	Q9EPK2	Q9EPK2 rattus norvegicus
15	1033.5	13.5	1061	4	Q13420	Q13420 homo sapien
16	545	7.1	416	4	Q13077	Q13077 homo sapien

17	508	6.7	203	4	Q9BR49	homo sapien
18	495.5	5.6	2035	5	Q9G519	Q9br49 trospophila
19	425	5.6	1778	5	Q9VLD8	Q9vld8 trospophila
20	374	4.9	326	4	Q15406	Q15406 homo saplen
21	340	4.5	2232	5	P91365	P91365 caenorhabdi
22	340	4.5	2302	5	Q9N653	Q9n653 trospophila
23	339.5	4.4	893	3	Q9N6J4	Q9n6j4 brachydanit
24	339	4.4	2310	5	Q9GR49	Q9gr49 trospophila
25	334	4.4	5476	5	Q9NJ17	Q9nj17 trospophila
26	334	4.4	5533	5	Q9VPL2	Q9vpl2 trospophila
27	334	4.4	5554	5	Q9NHN1	Q9nhn1 trospophila
28	334	4.4	5560	5	Q9VPL1	Q9vpl1 trospophila
29	332	4.4	2280	5	Q9V8E6	Q9v8e6 trospophila
30	328	4.3	862	11	Q9WVS9	Q9wvs9 ratius norvi
31	328	4.3	865	11	Q9IYB0	Q9iyb0 spalax galli
32	328	4.3	5533	5	Q9UC3	Q9uc63 dtrosophila
33	325	4.3	865	11	Q9IYB2	Q9iyb2 spalax carm
34	325	4.3	865	11	Q9IYB8	Q9iyb8 spalax juda
35	321.5	4.2	778	13	Q9PNA3	Q9pna3 xenopus lae
36	321	4.2	825	13	Q9IY06	Q9iy06 xenopus lae
37	319	4.2	875	13	Q9W7C3	Q9w7c3 gallus galli
38	318	4.2	853	13	Q9W6G2	Q9w6g2 gallus galli
39	314.5	4.1	870	6	Q9XTA4	Q9xta4 bos taurus
40	314.5	4.1	2703	5	Q9VE87	Q9ve87 dtrosophila
41	313.5	4.1	4957	4	Q14687	Q14687 homo saplen
42	313.5	4.1	5262	4	Q14686	Q14686 homo saplen
43	312.5	4.1	832	11	Q920Y1	Q920y1 ratius norvi
44	310.5	4.1	852	13	Q9IYK1	Q9iyl1 gallus galli
45	309	4.0	2715	5	Q61603	Q61603 dtrosophila

## ALIGNMENTS

RESULT	1	
098RTW1		
ID	098TW1	PRELIMINARY;
AC	098TW1	PRT; 1505 AA.
DT	01-JUN-2001 (TREMBLrel. 17, Created)	
DT	01-JUN-2001 (TREMBLrel. 17, Last sequence update)	
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)	
DE	BRANCHYPTERIAL INTERMEDIARY FACTOR 2.	
OS	Brachydanio rerio (zebrafish) (zebra danio).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;	
OC	Cypriniformes; Cyprinidae; Danio.	
OX	NCBI_Taxid=7955;	
XX	[1]	
RN	SEQUENCE FROM N.A.	
RA	Tan J.H., Kara A.U., Chan W.K.;	
RP		
RT	"Molecular cloning and characterization of zebrafish nuclear receptor	
RL	coactivator 1TF2.";	
RL	Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.	
DR	EMBL; AF323989; AAK1608.1;	
DR	InterPro; IPR001092; HLM.dlm.	
DR	InterPro; IPR000047; PAS.	
DR	Pfam; PF00989; PAS. 1.	
DR	SMART; SM00353; HLM. 1.	
DR	SMART; SM00091; PAS. 1.	
SO	SEQUENCE 1505 AA; 161793 MW; 888DDA8435B14497 CRC64;	

Query Match	55.7%;	Score 4248;	DB 13;	Length 1505;
Best Local Similarity	58.3%;	Pred. No. 3.9e-251;		
Matches 900;	Conservative 129;	Mismatches 310;	Indels 154;	Gaps 36

[illegible]



```

Db 178 VNGVSMNTETOROKSHTEFCRMIMKT PHDILEDINASPEMRORYETMOCFALSOPRAME 237
Qy 241 BEEDLOSCLICARVAPKPERVLPSSSEFTTRDLOGKITSLDSTMAANKPGMEDIV 300
Db 238 BEEDLOSCLICARVAPKPERVLPSSSEFTTRDLOGKITSLDSTMAANKPGMEDIV 297
Qy 301 RRCICRFAHOGESVYAKRHHHEVLROGLAFSOIYFRSLSDGLVAOQTSKLIRSOY 360
Db 298 RRCICRFAHOGESVYAKRHHHEVLROGLAFSOIYFRSLSDGLVAOQTSKLIRSOY 356
Qy 361 TNEPOLVLSLMLHREOVNCPVNDLGTQMGKPLNPISNSPAHQALCSGPGQDMTIS 420
Db 357 TNEPOLVLSLMLHREOVNCPVNDLGTQMGKPLNPISNSPAHQALCSGPGQDMTIS 412
Qy 421 SNINPINKPKRQMGMPRGSGGGMNHS---GMOA--TFPGSGNYALKMNSPSSSGC 476
Db 413 SSRATGLADPTTGGMGCARITGSSNINSLTTPGQMOSSTYONNNGYGMSSPPHSGTG 472
Qy 477 MNPQOPTSLSPRHMSRPGVAGSPRIIPSPQSPAGSLSPVGVCSSTGNSHSTYNSLNA 536
Db 473 LAPNOQNMISPRNR-----GSPKIASHQFSPVAGVHSPMASGNTGN-HSFSSSLSA 525
Qy 537 LQALSEGCVSLGSSLASPEDLKMGILONSVPVMMNPPLSKGSLDKDCFLYGE--PSE 594
Db 526 LQALSEGCVSLGSSLASPEDLKMGILONSVPVMMNPPLSKGSLDKDCFLYGE--PSE 581
Qy 595 GTTGAESSCHPEQKETNDPRLPRAVSEERADGOSRLHDSGQTKLLOLITTKSD--- 650
Db 582 SSMCOSNSRDHLS-KEKES-----SYEGANORGPLESQGHKKLLQDLTLCSSDRGH 634
Qy 651 -OMEPSPLASL-----SDTNKDSGTSLPGSGHGTSLKREKHLIRHLQDSSS 699
Db 635 SSLTNSPLDSSCKESSVATSPSGVSSSTSGVSSSTNMHGSLLEKHLIRHLQDSSS 694
Qy 700 PYDLAKLAEATGKLSQESSSTAGSEVITQOEPRYSKKE-NALLRILDKDQTDIG 758
Db 695 PAEVAKITAATGKID---TSSITSCGDGVNVEQELSPKKKNALLRILDRDPSDL 751
Qy 759 LPEIPKLERLDSTDPANFTKLIAMKTEKE-EMSEFGDQDQSELDNEITLDDLONSQ 817
Db 752 SKELDPQVEGVNKMOSCTSTIPSSQKDKITETSESGDIDNDLIGDITSD 811
Qy 818 LPOLEPDRPGAPAGSVDKOALINDLMOLTAENSPVTPVGAOKTALRLSTFNNPQ 877
Db 812 FYNNSISSN-GSHLGT--KQGVFG-----GTSLSLGKSSQSV-QSIRP-- 850
Qy 878 LGRLLPNONPLD--ITLQSPGAGPPPIRNSPYVIP-OPGMNGOMIGNOGILGN 934
Db 851 -----PYNRAVSLDSPVSVGSSPVKNISAFPLPKQPMIGNPRMDOSENGS 900
Qy 935 SSTGMIGSASRPMPSGEMA-POSSAVRYTCAATTS-----AMRPVQGMIRN 983
Db 901 SMGGRNRYVYTYOTSSDGMGLPNSKAGMEPMNSNMGREGDYNTSLPRALGSI-- 958
Qy 984 PAASIPMRSSPOGQOTL--OSQVMNIGPSELENNMGPOYSQOQAPNPQAPPESTL 1041
Db 959 --PTLPKLSNISPARGVYLQOQOQMLQMRGEIIPMGKANPYG-QAASNOIGSPDGL 1015
Qy 1042 PILDASFSQNRQPGSSPDLLCPHPAESPDGALLDLYLALRNF-DLEIDRA 1099
Db 1016 SMDVSHQTORPRLRNLSDLVGPPSNLEGGSDERALLDHTLSTNDATGEIDRA 1075
Qy 1100 LGIPELVASOAVDEPO--FSSQDSNIMLEOKAPVFPQOYASQAQAGSVSPMDPNH 1157
Db 1076 LGIPELVASOAVDEPO--FSSQDSNIMLEOKAPVFPQOYASQAQAGSVSPMDPNH 1127
Qy 1158 TMGQPSYATLRMQ-----PRGLRPTGLVQ---NOPNOLRLQLOHRLQAO---ON 1202
Db 1128 LOGGSPFSNMNMGNOGQNPLOGMHPRANIMRPTMPPKRLMQLOQRLOGQOFLNGS 1187
Qy 1203 RQPLMNQJLN--VSNNVNLRLRQV--PTQAPINAQMLAQORELNLNHLR----- 1248
Db 1188 RQALETKMENPTAGAAVVRPMQPOGFLNAQVARSRELLSHHFROQVAMMMQOQ 1247

```

```

Qy 1249 -----OROMHOQOOVORTLMMRGQGLNMTSPVAPSGMDATMSNPRIPOANQOPPE 1301
Db 1248 QOOOQOOOQOOOQOOOQOOOQOOOQOAFSPPPNVTASMSGLLAGPMPAPPOQFPY 1307
Qy 1302 PPNYGISQOPDPFGTATTPOSPLSPMAHATOSPMMQOQANPAPVAPSDINGNAQGNM 1361
Db 1308 QNYMGQOQPDPAFQVSVSPPMAMMSRMGSPQNPMPMHQPAASITQS--SEKKGMPSGUL 1366
Qy 1362 GGNMFSQOQSPPHFGQOANTSMYNNMNTINVSMAINTGGMSSNMOTGOISMTSVST 1421
Db 1367 ARNNSFSQO---QFAHQGNPAVYS-----MVMHNGSSGHGMGNMNPMD 1408
Qy 1422 SGLSSMPEQ 1431
Db 1409 SGH-PMGPDQ 1417

RESULT 3
Q9UPG7 PRELIMINARY: PRT: 1417 AA.
ID Q9UPG7.
AC Q9UPG7.
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE RECEPTOR-ASSOCIATED COACTIVATOR 3.
GN RAC3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP MEDLINE=97385128; PubMed=9238002;
RA Li H., Gomes P.J., Chen J.D.;
RT "RAC3, a steroid/nuclear receptor-associated coactivator that is
RT related to SRC-1 and TRP2.*";
RL Proc. Natl. Acad. Sci. U.S.A. 94:8479-8484(1997).
DR EMBL: AF010227; AAC51663.1;
DR InterPro: IPR001092; HLA_dim.
DR InterPro: IPR000014; PAS.
DR Pfam: PF00989; PAS; 1.
DR SMART: SM00353; PAS; 1.
DR SMART: SM00091; PAS; 1.
SQ SEQUENCE 1417 AA; 15453 MW; 69550F581CB51B3 CRC64;

Query Match 35.8%; Score 2734.5; DB 4; Length 1417;
Best Local Similarity 42.8%; Pred. No. 1.4e-158;
Matches 645; Conservative 240; Mismatches 453; Indels 169; Gaps 44;
Qy 1 MSGCEBNTSDPSRAETRRKKECPDQGPSKNTKRNKEQENKYIEELIETIANFNDI 60
Db 1 MSGCEBNTSDPSRAETRRKKECPDQGPSKNTKRNKEQENKYIEELIETIANFNDI 59
Qy 61 DNFNPKPKCALTKETVROIKROKRAAANIDEVOKSDVSGGQVYIDDALGPMML 120
Db 60 DNFNPKPKCALTKETVROIKROKRAAANIDEVOKSDVSGGQVYIDDALGPMML 118
Qy 121 EALDGEFFVNVLEGNVVESENVTOYLRYNOBELANKSVYSILYAGDHTFEYKNLIPRSI 180
Db 119 QALDGEFLVNVNEANIVFSENVTOYLOKQDVAINTSVYNNLHREDRDKFLKN-LPST 177
Qy 181 VNGSMSGSEPPRNRSHTRCMLVAPLPDSEEGHNDQANOKYETMOCFAVSQKSIKE 240
Db 178 VNGVSMNTETOROKSHTEFCRMIMKT PHDILEDINASPEMRORYETMOCFALSOPRAME 237
Qy 241 BEEDLOSCLICARVAPKPERVLPSSSEFTTRDLOGKITSLDSTMAANKPGMEDIV 300
Db 238 BEEDLOSCLICARVAPKPERVLPSSSEFTTRDLOGKITSLDSTMAANKPGMEDIV 297
Qy 301 RRCICRFAHOGESVYAKRHHHEVLROGLAFSOIYFRSLSDGLVAOQTSKLIRSOY 360

```

```

Db 298 RCIORFFSINDGQMS-QRRHYOEAFLNGHAETPVYRESLADGTTVTAOTKCKLEPNV 356
QY 361 TNEPOLVISHMLHREQNVCMNPDLTGOTMGRPLNFISSNSPAHQALCSGNQODMTLS 420
Db 357 TNRHGVSTHFLQRENGVRRPNPNVYGQIRPRMAGCNSVGG----NMSPNQGLOMP 412
QY 421 SNINFPINGKEQMGMPMGFRGSGGNNHVS---GMOA-TTPQGSNYALKMNSPSSSPG 476
Db 413 SSRAYGLADSTTGOMSGARYGSSNIALTPPGMQSPSSYONNNVGLMSSPSPG 472
QY 477 MNGOPRMSLSPRHMSGAVGASPRTPPQSPAGSLHSVGVCCSSGNSHTNSSINA 536
Db 473 LAPNOONIMTSPKRR-----GSKTASHQFSVAGVHSMASGNGN-HSFSSSLSA 525
QY 537 LQALSEGHVGLSSLASPLDKMGNLQNSPVNNPPLSMGSLDSKDFGLXGE--PSE 594
Db 526 LQALISEGVGLSTLSTSSPGK---LDNSP-NNNITQPSKVSNOQDSKSPGLFCODQPVE 581
QY 595 GTTGOAESCHPEQOKETNDPNLEPPAVSSERADQSKRLHDSKQTKLLQTLTKSD--- 650
Db 582 SSMQCSNRDHLSD-KESKES-----SYEGAENQRGPLESKGHKLLQTLTSSDDRGH 634
QY 651 -QMEPSPLASL-----SDTNKDSVGLSPGSGTHTGSLKEKHKLTHRLDQSSS 639
Db 635 SSLTNSPLDSSCKESSVSVTPSGVSSSTSGVSTSNMHSGLDQEKHRLHKLQNGNS 694
QY 700 PVDLAKLTATGCKDLSESSSTAPGSEVYTKOEPVSPKKE-NALLRYLLDKDPTKDG 758
Db 695 PAEVAKTTAATGKD---TSTITSCGGGVNVKQEQSLPKKENALLRYLLDNDPDSAL 751
QY 759 LPEITPKLERLSDKTDPASNTKLIAMTEKE-EMSEFGDQPSGLDNLLEITDLONSQ 817
Db 752 SKELQPOVEVDNMKMSQCTSTTIPSSQEKDPKIKETSESGSDDLNLAITLDTLSSD 811
QY 818 LPQLFPTRGAGAGSDVKAIIINDLMQLAENSPTVPAQKTAIRLISOSTFNPNRPGQ 877
Db 812 FYNNSSISN-GSHLGT-KOOVFO-----GTNSLGKSSOSV-QSIRP- 850
QY 878 LGRLLPNQNLPLD--ITLQSPTAGPFPPIRNSSPYSVIP-QPGMNGNQMIGNOGMLN 934
Db 851 -----PYNRVSLDPSVSGSSPYKNTSAFPMILKQPKLGTGNGPRMMSQENYGS 900
QY 935 SSTGMIGNSASRPMPGSEMA-POSSAVRYTCAATTS-----AMNRPVOGGMIRN 983
Db 901 SMGGPNNTVYTOTPSSGDMGLPNSKAGRMPEMNSNMGPRGDDYNTSLRPAAGS1- 958
QY 984 PASIIPRPSSOPGQRTL--OSOVNIGPSELNMNGSGQYOSQQAAPRNOTAPWPSIL 1041
Db 959 --PPLPLRNSSTIPGARVYLOOQOQOMLOMRPELPMGANGPYG-QAASNQLSGWPDMGL 1015
QY 1042 PIDOASFASONROPFSSPDLLCPHPAESPSDEGALLDQLYLALNFD--GLEETDRA 1099
Db 1016 SMEVSHGTCQNRPLRNSLDDIVGPRNSLGEQSERALLDQHLHTLSNTATAGEIEDRA 1075
QY 1100 LGITELVYSQOAVDPPO--FSSQDSNIMLQKAPVPPQYASQAQMAQSGYSTPWQDNEH 1157
Db 1076 LGIPELVNQOALEPKODAFQGOEAAVMQKAGLYGQYPAOGPPMOG-----GFH 1127
QY 1158 TMOGRPYATILRMO-----PRGLRPTGLYO---NONNOLRLQLOHLQAO---ON 1202
Db 1128 LOGOSPFNSMAMONQOGNFPLOGMHRANIMRPTNTKOLRMQLOQRLOGOQFLNQS 1187
QY 1203 ROPILNOISN-VSNVNLTLRPV-PTQAPINAOMLAOREEILNOHLR----- 1248
Db 1188 RQALTELKMEPRTAGGAVMHRPMQPOGFLNAQMVARSSELSSHHRQQRVAMMMQOQ 1247
QY 1249 -----QROMHQOQVOQFTLLMRGQGLMPTPSMVAAPSGMPRTMSPRIPOANAQOPPPPN 1304
Db 1248 QOQOQOQOQOQOQOQOQOQOQOQOQAFGPBPNTVYASPMDDLAGPTPOAPPOOPYOPN 1307
QY 1305 YG1SQOPDPGFTGATTPQSPMLSPMAHNTQSPMMQOQOAPVQAPQADINDINGMAQNGN 1364
Db 1308 YGMOQDPDPAFGKVVSPFNAMMSRMGPSONPMQHPQOASITQY-SMKGMPSGNLARN 1366

```

```

QY 1365 SMFSQSPPHFGQOANTSMYSNNMNTNVSMATNTGMSNMONTGQISMTSVTSTSGU 1424
Db 1367 SSFSQO---QFAHQNPAYYS-----MVHNGSSGHMGQNMNPMPSMG 1408
QY 1425 SSMGPEQ 1431
Db 1409 -PMGPDQ 1414

RESULT 4
QY609 PRELIMINARY; PRT; 1424 AA.
ID QY609;
AC QY609;
DT 01-NOV-1999 (Tremblrel, 12, Created)
DT 01-NOV-1999 (Tremblrel, 12, last sequence update)
DT 01-JUN-2001 (Tremblrel, 17, last annotation update)
DE THYROID HORMONE RECEPTOR ACTIVATOR MOLECULE (Dti04G16.2) (NUCLEAR
DE RECEPTOR COACTIVATOR 3 (THYROID HORMONE RECEPTOR ACTIVATOR MOLECULE
DE TRAM-1, RECEPTOR-ASSOCIATED COACTIVATOR RAC3, AMPLIFIED IN BREAST
DE CANCER AIB1, ACTR)).
GN TRAM-1 OR NCOA3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98010595; PubMed=9346901;
RA Takeshita A., Cardona G.R., Kolbuchi N., Suen C.S., Chin W.W.;
RT "TRAM-1, A novel 160-kDa thyroid hormone receptor activator molecule,
RT exhibits distinct properties from steroid receptor coactivator-1.";
RL J. Biol. Chem. 272:27629-27634(1997).
RN [2]
RP SEQUENCE FROM N.A.
RA Skuce C.;
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF016031; AAC51849.1; -.
DR EMBL; AL034418; CAC17693.1; -.
DR InterPro: IPR001092; HLH_dlm.
DR InterPro: IPR000014; PAS.
DR Pfam: PF00989; PAS.1.
DR SMART; SM00353; HLH; 1.
DR SMART; SM00091; PAS; 1.
KW Receptor.
SQ
SEQUENCE 1424 AA; 155293 MW; 732CDF0423161679 CRC64;

Query Match 35.8%; Score 2729; DB 4; Length 1424;
Best Local Similarity 42.5%; Pred. No. 3e-158;
Matches 644; Conservative 243; Mismatches 451; Indels 176; Gaps 44;

QY 1 MSGMGENTSDPSAEFRKRKECPDQIGSPKRMTEKRNROEKYIELEALIFANPNDI 60
Db 1 MSGIGENL-DPLASDSKRLKLPDTPGQGLTSGEKRRQESKYIELEALISANLSDI 59
QY 61 DNEFPRDKCAILKETVQIQRIKEQEKAAANIDEVQKSDSVSTGGVYIDKDALGPMML 120
Db 60 DNRNVKRDKCAILKETVQIRQIRKEQK-RTSDDVDQKADVSTGGVYIDKXSLPILL 118
QY 121 EALDGEFFVNLGNVYVNSENTQYIRYNOELANKSVYSLHVDGHTFEVANKLPKSI 180
Db 119 QALDGEFFVNLGNVYVNSENTQYIRYNOELANKSVYSLHVDGHTFEVANKLPKSI 177
QY 181 VNGSWSGEPBRNSHTFENRMLVYKLPDSEEGHNOEAKHOKYETMOCFAVSQPSIKE 240
Db 178 VNGSWTNEQROKSHFNFRCMLKTPRHDLIEDINSPENRQRETMOQCFALSPRYAME 237
QY 241 EGEDLSCLICVARRVPMKEPVLPSSESTTTRQDLOGKITSLDTSTRAAMPGEEDLV 300
Db 238 EGEDLSQCMICVARRITTGRTTPSNPESRITTHDLISGKVVNIDTISRSMSMPGEEDII 297
QY 301 RCIORFHAOHEGESVYAKRHHHEVLRQGLAFSQTIRFSLSDGTLVAOTKCKLIRSGT 360

```



[illegible]

Query Match	Best Local Similarity	Score	DB #	Length
Matches 642; Conservative 242; Mismatches 439; Indels 194; Gaps	35.5%;	2711;	DB 4;	1412;
1 MSGCEMTSDSPRAETRRKRECEPDGLGSPKRNTEKRRNREDENKYIEELAEIIFANFNDI	1 MSGLEEDN-DPLASDSRRKKLPCCDPPGGGLTCSGEKRRREDESKYIEELAEIISANLSDI	60		
1 MSGCEMTSDSPRAETRRKRECEPDGLGSPKRNTEKRRNREDENKYIEELAEIIFANFNDI	1 MSGLEEDN-DPLASDSRRKKLPCCDPPGGGLTCSGEKRRREDESKYIEELAEIISANLSDI	59		
61 DNFNPKPKCALIKETVKQIROIKEOEKAAANIDEVOKSDVSSNGOGVIRKDALGPMWL	60 DNFNPKPKCALIKETVKQIROIKEOEKAAANIDEVOKSDVSSNGOGVIRKDALGPMWL	120		
121 EALDGFEEVFNLEGNVVESENVTYLRNOBELANKSVSYLHVGDTIEFYKNIPLPKSI	119 QALDGFEEVFNLEGNVVESENVTYLRNOBELANKSVSYLHVGDTIEFYKNIPLPKSI	180		
181 VNGSGWGEPPRRNSHTNCRMLVLPDLPSEEBGHNDNOAHOKRYETMOCFAVSQPKSIKE	178 VNGSGWGEPPRRNSHTNCRMLVLPDLPSEEBGHNDNOAHOKRYETMOCFAVSQPKSIKE	240		
241 BEEDLQSCILICARVPMKPERVLPSSBSFTTRDLODKITSLDTSIRAMKPGEMDLY	238 BEEDLQSCILICARVPMKPERVLPSSBSFTTRDLODKITSLDTSIRAMKPGEMDLY	300		
301 RCIQCFKFAHGECEVSYAKRHHHEVLKQGL-----AFSQIYFFSLSDGTLYAAQ	298 RCIQCFKFAHGECEVSYAKRHHHEVLKQGL-----AFSQIYFFSLSDGTLYAAQ	350		
351 TKSILIRQTTNEPOLVLSLHMLHREQNVYVNPDLTGOTMGKPLNPISSNSPAHQALCS	349 TKSILIRQTTNEPOLVLSLHMLHREQNVYVNPDLTGOTMGKPLNPISSNSPAHQALCS	410		

```

Db      1352 GMPDGLNANSSFSQO---QFAHOGNPAVYS-----MHMNGSGHMGOM 1393
Qy      1415 SVTSYSTGLSSMGPEO 1431
Db      1394 MNMMPMNSGM-PMGPDQ 1409

RESULT 6
057539 PRELIMINARY: PRT: 1391 AA.
AC 057539:
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE RETINOID X RECEPTOR-INTERACTING COACTIVATOR XSRC-3.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Piploidea; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxId=8355;
RN [1]
RP SEQUENCE FROM N.A.
RA Kim H.-Y., Lee S.-K., Na S.-Y., Choi H.-S., Lee J.W.;
RT "Molecular cloning of xSRC-3, a novel retinoid x receptor-interacting
RT coactivator from Xenopus, that is related to AIB1, p/CIP and TIF2."
RL Mol. Endocrinol. 0:0-0(1998).
DR EMBL: AF044080; AAC12927.1; -.
DR Interpro: IPR001092; HLH.dlm.
DR Interpro: IPR000014; PAS.
DR Pfam: PF00989; PAS; 1.
DR SMART: SM00353; HLH; 1.
DR SMART: SM00091; PAS; 1.
DR Receptor.
KW SEQUENCE 1391 AA; 152532 MW; AD28F5CD934AC33D CRC64;
SO

Query Match 35.4%; Score 2698.5; DB 13; Length 1391;
Best Local Similarity 43.3%; Pred. No. 2.1e-156;
Matches 648; Conservative 239; Mismatches 436; Indels 173; Gaps 51;

Qy      1 MSGGENTSDPSAETRRKKR--ECPPDLGSPPRMKEKRNREOKNYIELAEILIPANFN 58
Db      1 MSGIGENSIDPLASETRKKRPSSC-DTPGGDLTGCSEKRRREDESKYIELADLIANIS 59
Qy      59 DIDNFKRDKCALIKETVQKIQIKOEKAAANIDEVQKSDVSSTGGGVIDKDALGPM 118
Db      60 DIDNFVVKRDKCALIKETVQKIQIKOEK--ASSNDVDQKADVSSTGGGVIDKDALGPM 118
Qy      119 MEALDGFPPVYVNLBEGNVYFVSENVQYLYRNOELMKNKSVSYILHVGCHTEYVKKLLRK 178
Db      119 LQALDGFYVYVNRBSIYFVSENVQYLYQKQEDLVNYSYILHBEEDRKDFLKN-LRK 177
Qy      179 STYNGSGWSEPPRRNSHFNCRLVYKPLPDSSEEGHNOEAKOKETMOCFVASFQPKSI 238
Db      178 STYNGVPMFSETPRKRSHFNCRLVYKPLPDSSEEGHNOEAKOKETMOCFVASFQPKSI 236
Qy      239 KEEGEDLOSCILICVARVPKERRVLPSSSEFTTRDLOGKITSLDTSTWRAAMKGEWED 298
Db      237 IEEGEDLOSCIMICVARITTAERAFSANPESFTTRDLDLGKVVINDANSIRSSMRGFEED 296
Qy      299 LVNRCTQKFAHOGESVSTAKRHHHEVLROGLAEFQIYRFSLSDDTGLVLAQTKSKLINS 358
Db      297 TIRRCIORFLFHSQGPWTY-KRHYDEAVYHGLSETPLYRFFSLADGTMYTAQTKSKLIFRN 355
Qy      359 QTNPEPOLVLSLMLHREOVNVCVANNPDILGOTGKPLNPITSSNPAHQALCSGNPDQDMT 418
Db      356 PVTNDHGFVSTHFLRDEQNGYRPANPMAQG:TRPOANPULPNT-----MNSMPQAMQ 409
Qy      419 LSSNINFLIPKPEQMGKMPGRFGSGGM---NHVSGMATTPQ-GSNYALKNNSPSSOS 474
Db      410 -QGNRNYGMGDPRNSMQMGMRKYSFGNMAPVYQAQGVQGSPPYQNNNSNTGLNNSPFRHS 468
Qy      475 PGMNPGOPTSMLSPRHRMSPGAVGSPRIIPSPQSPAGSLHSPVGVCSSTGNS--HSTYNS 532
```

```

Db 469 PGMANQPNLWVSPRNRASPKMAS-----NOFSVPGMNSPMG---SSGNAGGSFSS 519
Oy 533 SLNALQALSEGHVSLGSLASPDLMKGNLONSPPVNMNPPRLSKGSLDSKCFGLXGP 592
Db 530 SLNALHAISEGSSLSLSSPGQKVEN--NS--NMNPMQKICNDCKSPSLGYCE- 574
Oy 593 SEGTLGOAESS-CHPEGEKETMDPMLPPAVSERADGOSRLDSKGOKRLDOLLTKSDQ 651
Db 575 ----OGVSSVCOSSGSEHLEKDYKENI-FEGSESQSAESGKHLKDLLOLCFTEE 629
Oy 652 MEPSPLASLSDTNKDSGLSPSG-----STGCTSLKEKHLRLDLODS 698
Db 630 RGSISLSSSSMDC-KDSNVNTPSGCVSSSTIGVSTSLHGSMLQEKHRLHLKLQNGN 688
Oy 699 SPVDLAKITAEATGKDLQESSSTAPGSEVTTKQEPVSPKKE-NALLRYLLDKDXTDI 757
Db 669 SPAEYAKITAEATGKDVQETVSSAPCTEATVKRQLSPKKENNALRLHLDLKDWMDP 748
Oy 758 GLPETPKLERLDSKTDPAASNTKLIAKTEKE-EMSFEPGDPGSELDNLEETLLONS 816
Db 749 LAKDKPVEHMDIKMGSCSSSNVPTSSODKEVKIKTEPGEVPGDLDNLDAIILDLAGS 808
Oy 817 QLPOLFPDTPRGAPAGSVKQAIINDLMQLAENSPTVPGAKTALRISQSTFNNPRPG 876
Db 809 ---DFYSNMSBRASDLQKQVFDSPPL-AMRSPDSMGSR----- 847
Oy 877 QLGRLPQNPLDITTLQSPGAGFPPIRKNSSPVTPQGMGMNQGMLGSLSS 936
Db 848 -----PPFNAMSLDSRST-----PPVNVNPFMLPKQMSLSPRMQODNFG- 893
Oy 937 TGMIGNASRP--TWSEGMARPOSSAV-----RYTCAAT-----TSAMNRPVQGMIRN 983
Db 894 -VMMSGNRKSNQHPGDMAMQNSAVNRLERPVGSGRGCRPDYSSATPRAMGGM- 950
Oy 984 PAASIPMRSSOPGQROTLDQSO--VMNIGPSELENNMGPOYSQQAAPPNOTAPRESIL 1041
Db 951 --PGLITRNSNIPGSRPVYMQOQNHILPWRPNDMAMSGSNPYG-QQASNPBGSPDAIM 1007
Oy 1042 PIDDASFASQNRKQRPFGSSPDLLCPHRAEBSDEGALLDLYLALRNF--GLEIDRA 1099
Db 1008 -MNGRGGAQNRKQRLSLDLCLCPSTVEGOTDEILDLLHTLSTYDAGLEIDRA 1066
Oy 1100 LGIPELVASQAVD--PROFSSODSNIMLEOKARVFPQOYASO-QOMOGSVSPMODNF 1156
Db 1067 LGIPELVASQALEPQPDSTORQSGRVMIQKPRMIGOHYAGQAMAGGNNMQ---- 1122
Oy 1157 HTMGORPSYATLR-----MORPGLRP--TGLVQ--NQPNOLQLOHRLDAQ---- 1200
Db 1123 ---GQHPFNIVMGOMNQOQGMHNPLOGMHRANLIRPNNNIRPKOIRMOLOQRLQOQFLN 1179
Oy 1201 QNRORLMAQISNVSNVNI--TLRPVYPT---QAPINAMLAOROREILINOLRQRMHQ 1255
Db 1180 QNRQALEKXVPMRPMGAGVAPVMOQTPVSOQGFNAQMVAKNREILISHQIRQMMAM 1239
Oy 1256 QOVQORTLMMGGQGLMTPRSVAVRSGMPTMSNPRIIPANQOQFFRPNTYGLSOQPDGF 1315
Db 1240 MQOQOG---OPQAFSPRPNTYASASMDNPLGGPRMPQAPPOQFSYPRNYGINTDPTF 1295
Oy 1316 TGAITPOSPLASPRMAHTQSPMOMOQANPAYQAPSDINGAQMNGNSMSPHF 1375
Db 1296 GRVSSPRAMSSRMAPSQNHPTQMYPS----PDKMGPSGMARPNSTPQOQYSH- 1350
Oy 1376 GQOANTSMTSNMNNINVSAMATNGMSSMNQMTGOISMTSVTSGLSSGAPQ 1431
Db 1351 --QTNPATY-MMHMHN-----GNGNHMGOM-----NINSLPMGSM-PMGPDQ 1388

```

```

Dt 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
Dt 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE P300/CBP/CO-INTEGRATOR PROTEIN.
CN NCOA3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9736097; PubMed=9192892;
RA Torchia J., Rose D.W., Inostroza J., Kamel Y., Westin S., Glass C.K.,
RA Rosenfeld M.G.;
RT "The transcriptional co-activator p/CIP binds CBP and mediates
RT nuclear-receptor function."
RL Nature 387:677-684 (1997).
RN [2]
RP SEQUENCE FROM N.A.
RA Torchia J., Rosenfeld M.G.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF000581; AAC05020.1; -.
DR TRANSFAC: T04638; -.
DR MGD: MGI:1276535; Ncoa3.
DR InterPro: IPR001092; HLH_dim.
DR InterPro: IPR000014; PAS.
DR Pfam: PF00989; PAS. 1
DR SMART: SM00353; HLH. 1.
DR SMART: SM00091; PAS. 1.
DR SMART: SM00091; PAS. 1.
SO SEQUENCE 1398 AA; 151573 MW; EF44E923735816C24 CRC64;

Query Match          32.9%; Score 2511; DB 11; Length 1398;
Best Local Similarity 41.2%; Pred. No. 6,4e+145;
Matches 621; Conservative 232; Mismatches 466; Indels 188; Gaps 47;

Oy 1 MSGMGENTSDPSRAETRRKKECPDQIGSPKRTKRNREQNKKYIEELAEILIFANFDI 60
Db 1 MSGLESSLIDPLAASRRKRLPCDAPGQGLVYSGEKKMRREQSKYIEELAEILSNLSDI 60
Oy 61 DNFNFKPKCALIKETVKQIROIKBOEKAANIDEVOKSDVSTGQGVIDKALGPMML 120
Db 61 DNFNFKPKCALIKETVKQIROIKBOEKAANIDEVOKSDVSTGQGVIDKALGPMML 119
Oy 121 EALDFFEVVNLGWNVYSENVTOYLKRYNOELMKNKSYSLHGHDTPEYKNLPSKI 180
Db 120 QALDFFLVVNDGNIVSENVTOYLQYKQEDLVNTSYSLHPRKDFLNTYQNDQL 179
Oy 181 VNGSGMSGEPRRNSHTNCRMLVRLPDSSEEGHNDQBAHQYETMOCFAVSOPKSIKE 240
Db 180 MEFGLMTRDKKAKYIILVRLMK--THDILEDVNASFETKORYETMOCFALSOPRAMLE 238
Oy 241 EGEDLOSGLICVARRVPKKEPVLPSSEFTTRDLOGKITSLDTSTKRAAKPMEDLV 300
Db 239 EGEDLOSGLICVARRV---TAPFPSPSPSEFTTRHDLGKVYVIDNLSRSSMRPFEDI 295
Oy 301 RRCIOKFAHQEGESVSTAKRHHNEVYLQGLAFSOTIYRFSLSDGTLVAQTKSKILRSOT 360
Db 296 RRCIORFESLNDGQWS--OKRHYQEAUVYGHAEFTVYFESLADGTIVSAQTKSKLFRPV 354
Oy 361 TNEPOLVLSLHLHREOVYVANNPDLTGOTMGPRLNPISNS---PRAHOLACSNPQD 416
Db 355 TNDRGFISTHFLQREON-----GYRNPPLIPDQGIKPPAAGCVASHPMN 401
Oy 417 MTLSSNINFPINGPKQGMKPGRGSGGNHVS--GQOATTP---QGSNTALKNMSPQ 472
Db 402 YQMGMSRTYGVDDPSNTQMGAGARVAGASSVASILPQGSLSQSPSYQSSYGLSMSPRH 461
Oy 473 SSPGMNPGOPTSMLSPRIRMSGVAGSPRIPPSQSPAGSLSPVGVSSSTGNSHTYNS 532
Db 462 GSPGLGPNQOONIMISPRNR-----GSPKMASSHQSPAGASHPGPGNTG-SHSFSS 514
Oy 533 SLNALQALSEGHVSLGSLASPDLMKGNLONSPPVNMNPPRLSKGSLDSKDCFCFLGYCE- 591

```



```

Db 681 SLEAERHKLHRLLOE-GSPSDITTLVSPEPKKD-SASTSVSVTGOVGNSSIKLELDAS 738
Qy 738 KKKK---NALLRYLLDKD-----TKDIGLPEITPKLERLSDKTPASNTKLIAKTEKE 789
Db 739 KKKESKHQRLRYLLDKREKDLRSTPNLSLDVAKYKKE-QMPCNTNPTPKTPPE 797
Qy 790 EMSEFPDQPSSELDNLEIIDLON-SQLPOLFPDTPRGAPAGVDKQAIINDLMQTLA 848
Db 798 EIKLEAQSOTFADLDQFQOLLPTLEKAQOLPGLCETDRMD--GAVTSVTIKSEL----- 849
Qy 849 ENSPTPYGAOKTALRISQSTFNNRPGQLRLPNQMLPIDITLQSTPGACFPPIRNS 908
Db 850 -----LPSAQSAT----- 858
Qy 909 SPYSVLPQGMNGMNGMIGNLSSTGMIGNSASRPTMPSGMAVQSSAVRYTCAAT 968
Db 859 -----ARP----- 861
Qy 969 TSANRPVQGMIRNPASIFMRSSQPGOROTLOSQVMNIGPSELEMMNGGPQYSSQQA 1028
Db 962 TSRLNR-----LP-----ELELALDNOFGQPGT 885
Qy 1029 PPNQATAMP-ESLIPDQASASQNRQPGSSPDLLCPHPAESPDGALLDOL--YL 1085
Db 886 --GDOIPTNNTVTAIINQ--SKSEDOCISSIDELLCPTTVEGRNDEKALLQELVSEFL 940
Qy 1086 ALRNPDLGEEIDRALGIPELYSQSOAVD--PEQSSQSDN--IMLEOKAPFPQOYASQA 1141
Db 941 SGKQETELAEIDRALGIDKLIV-QGGGLDVLSERPQOATPPLIMEEPNLYSOPYSAPS 999
Qy 1142 QMAQSYSPMODPNFHTGQRPSTATLMOPR-----GLRPTGLYON--QPNOL 1189
Db 1000 PTAN-----LSPRQGMVWRQKPSICTMVOYTPRGARSPGMQNPRTILNRPAPNOL 1054
Qy 1190 RLQLOHRLQAO-----QNRQPLANOISVSNVNTLTPG--VETQAPINAOMLAQOR 1240
Db 1055 RLQLOQRLOGQOOLIHNRQAILNOFATAPVGIIMRSGMOQOLTPPQPLNAOMLAQOR 1114
Qy 1241 EILNQHRLRORMHQOQOYQOATLMMRQO--GLNMPMSVAVASGMAPATMSNRIIPANNO 1298
Db 1115 ELYSQOHRORLQO-----QQRAMLRQOQSGNNLPPS--SGLEVQGNPRLPQAGADQ 1166
Qy 1299 FPPPNYGISQOQPPGFTG-----ATTPQSPV-----MSPMA 1331
Db 1167 FPPPNYGTNCTPASTSPESQLANPEASLANRNSVSGMTGNIGQGTGINTPQMO 1226
Qy 1332 HT-----QSPMAQSOANPAYQAPSDINGW 1356
Db 1227 QNVEQYPGAGVPOGEANFAPSLSPGSSMVPMPPIPPQSSILQOTPPASGYQSP-DMKAW 1285
Qy 1357 AOGMMGMSRQSO--QSPPHGQOQANTSMYNNMNNINVMATNTGMSMMQNTGQISMT 1414
Db 1386 QOQAGINNANVSAQVONP--TPAQPGVY-NNMSTVSMAGNTNVQNMNPMMAQOM- 1340
Qy 1415 SVTSVSTGSSMSGPEQVNDPALKGNLFPNQLPGMDIMIKQECDTTR 1462
Db 1341 --SSLQMPGMNTVCPQOINDPALNHTGLYCQNLSTDLKTEADCTQ 1386

```

```

RESULT 9
ID 043793 PRELIMINARY; PRT: 1399 AA.
AC 043793:
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE STEROID RECEPTOR COACTIVATOR 1E.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A.

```

```

RX MEDLINE=98090466; PubMed=9427757;
RA Kalkhoven E., Valentine J.E., Heery D.M., Parker M.G.;
RT "Isoforms of steroid receptor coactivator 1 differ in their ability to
RT potentiate transcription by the oestrogen receptor.";
RL EMBL J.17:232-243(1998).
DR EMBL; AJ000882; CAA04372.1; -.
KW Receptor.
FT CHAIN 2 1399 POTENTIAL.
SQ SEQUENCE 1399 AA; 152384 MW; DACE967BJ3AC6B69 CRC64;

Query Match 28.5%; Score 2174; DB 4; Length 1399;
Best Local Similarity 36.2%; Pred. No. 2,6e-124;
Matches 581; Conservative 236; Mismatches 422; Indels 368; Gaps 61;

Qy 1 MSGCEMTSDPSRATRRKKECP-DQLQPSKRNTEKRNDEQNKYIELLALIFANRD 59
Db 1 MSGLDSSSDPANPDSHKRGSPCDTLAS--TEKRRRQENKYIELLALISANISD 56
Qy 60 IDNFKRDKALKKETKQIRQIKQEKAAANIDEVOKSPVSTGGVIDKDALGPM 119
Db 57 IDLSLVKPKDKILKTYDQIQMKRMEQEKSTDDQVOKSDISSQGVIEKESLGPL 116
Qy 120 LEALDGEFFVNLGENVVFESENQYLRVNOELMNKSVSILHVGDTFEVKMLPKPS 179
Db 117 LEALDGEFFVNCBGRIFYVSENTSYLGYNQELMNTSVSILHVGDAEFVKKMLPKS 176
Qy 180 IVNGSGSGEPRRNSHTFNCRLMVLPLPDSEEGHNOEAHQKTYMQCAVSOQPSIK 239
Db 177 LVNGVPMPQEAETRRNSHTFNCRLMILHP-PD--EPGTMEACQREVWQCFVSOQPSIK 233
Qy 240 EGEDQSCICLVARVYMKERPVLPSSSEFTTRDDLOGKITSLDTSTMRAMKGMWDL 299
Db 234 EGEDQSCICLIARLP--RPALTGVSEFTKQDITTKIISIDTSSLRAGRTGWBEL 291
Qy 300 VRCIOKFAHQEGESVYAKRNHHEVLROGLAFQIFRSIDSTLVAAQTKSKLIRSQ 359
Db 292 VRKCIYAF--QPGREPRYARQLPQEVMTRGASSPSYRFLINDQTMLSAHTKCLCPQ 350
Qy 360 TTNEPOLYISLMLHREONVCMNP-DLTGOTMKP-----LNPISNSPAHQALCSN- 412
Db 351 SPDMQPFITMGIIIDREHS--GLSPQDDTNSGMSIPRYNPVNP--SLSPAGVARSSTL 406
Qy 413 -PGQDMTSSNINFPINPKED-----KMPMGRCGGSGMNHVSGM-----QA 455
Db 407 PPSNMMVSTRIN-----RQSSDLHSSHSNNSNSGFCSPGSOIVANVALNGQA 460
Qy 456 TTPQGSNTALKMNSPQSSQPGMNPQPTSMLSPPRRHMPGVAGSPRIPPOFSP-AGSLH 514
Db 461 SS-QSSNPSLNLNNSPMTGTGSLAQ---FMSPRQVYSGIATRMRMNNSEFPPISTLS 516
Qy 515 SPVGYCSS--TGNSHYSTNSSLNALQALSEGHVSLGSSLASPDLMKNTLONSPVNMNP 572
Db 517 SPVGMTSSACNNNNNSYINIPYSLQMGNEGPNNNSVGFSASAPYLQKSSONSPRLNIQ 576
Qy 573 PLKMKGISDSDCRGELYEPSEGTGQAESSCHPEQKET---NDPMLPRAVSSERADG 628
Db 577 P-AAESKDNKEIASILNEMI-----QSDNSSQDKPLDLSGLLHNNDL-----SDG 622
Qy 629 QSLHDSKQTKLQLLQTLTKSD-QMEPSPILASSLD-----TKKDTGSLPGSGSTHG 680
Db 623 DSKY--SQTSHKLVOLLTTTAEOQLRHADITPSCADVLSCTGTSASANSAGSGSPSH 680
Qy 681 TSLKERRKILHRLLODSSSPVDLAKLTAEATGKDLISQSSSTAPB--SEVTIKQEPVS 737
Db 681 SLEAERHKLHRLLOE-GSPSDITTLVSPEPKKD-SASTSVSVTGOVGNSSIKLELDAS 738
Qy 738 KKKK---NALLRYLLDKD-----TKDIGLPEITPKLERLSDKTPASNTKLIAKTEKE 789
Db 739 KKKESKHQRLRYLLDKREKDLRSTPNLSLDVAKYKKE-QMPCNTNPTPKTPPE 797
Qy 790 EMSEFPDQPSSELDNLEIIDLON-SQLPOLFPDTPRGAPAGVDKQAIINDLMQTLA 848

```







```

Db 690 RHKLHLRLOE-GSPSDITTLTSLVEPEKDSVPASTAVSOGSGASIKLELDAARKE 748
Oy 742 ---NALLRYLLDKD-----TKDGLPEITPKLERLDSKTPASNTKLIAMKTEKESESF 793
Db 749 SKHQHLLRYLLDKREKDLRSTPNCILDVKKYKKE- QMDPCNTNTPMTKRAPEEYKL 807
Oy 794 EPDGPSELNLEILLDION-SQLPOLF-PDTRPGAPAGSYDKAIIINDLMQLTAEENS 851
Db 808 ESQGFADLDQFOLLPLEKAQPLSLCETBDMCAVTC-----VSIRKEA-- 854
Oy 852 PYTVVGAQKALRISOSTFFNNRPGQGRLLPNQNLPLDITLQSPGAGPPPIRNSPY 911
Db 855 -VTPASLQPTTARA--PR--LSRL-----PELEBAID 883
Oy 912 SVIPQPMGNOGMIGNOGNLNNGSTGMIGNSAPTPMPSGEMAPQSSAVRVCATTSYA 971
Db 884 NOGQPG-----AGDQ-----IP--WANN--LTTIN 906
Oy 972 MNRPVQGMIRNPAASIPMRPSQPGORQLTQSVNIGSELEMMNGQDQYSSQQAAPPN 1031
Db 907 QNKP-----EDQCISQL----- 919
Oy 1032 QTAWPESILPIDIASFASQNRQPGSSPDILCPHAPAESPEGALLDOL--YLALRN 1089
Db 920 -----DELCPPTVEGRNDEKALLEQVLSFGKD 950
Oy 1090 FDGLEIDRALGPELVYSQSAVD--PEOFSODSN--IMLEOKAPFPOQYASQAQMAQ 1145
Db 951 ETELAELDRALGIDKLV-QGGGLVLSERPPQATPPLIMEDRPYLISQYSSPSPLA- 1008
Oy 1146 GSYSPMODPNTMTGORSYATLMO-----PRGLRPTGLYON--QPNQLRLQL 1193
Db 1009 GLSGPFG-----AVRKRPSLIGAMPVQYTPRGTSFPMNMGPROTLNRPAPAPNQLRLQL 1064
Oy 1194 QHRLOAQ-----QNRPLMANISVSNVNLTLRPG-----VTOAPINQMLAQORELIN 1244
Db 1065 QORLOQGOOLMHNQKAILNQPANAPRGVNMRSQMGOQITPOPPLNAQMLAQORELYS 1124
Oy 1245 QHLRORHQQOQVQOATFLMARGO--GLNTPPSWAPASGMPATMSPRIIPANQOPEFP 1302
Db 1125 QOHQOROIIO-----QORAMLMRHSFGNNIPPS-----SGLFVOMGTRRLPGAGPOOFPYF 1176
Oy 1303 PNYG-----IS 1308
Db 1177 PNYGTNCTPRASTPSFSQLANPEASLATRSSWVNRGMAGMNGQGCAGISPMQOVVF 1236
Oy 1309 QQPRPG-----FTGATTPQSPMLSPRAHNTQSPMMQOQANPAYQAPBDINGMAQGN 1360
Db 1237 QYPPGLVPOGEATFAPLSLFGSSMVPVPYPPQSSLLQOTPTPSYQSP--DMKAWOQGT 1295
Oy 1361 MGSMSFQ--QSPPHFGQOQNTSMYSNNININYSMATNTGMSMNOQNTQISMTSVTS 1418
Db 1296 MGNNTVSQAQVQSP--APRQPGVY--NMSITVSMAGNANNINQNMNPMQOMQOM---SS 1348
Oy 1419 VSTGLSGMPEQVNDPALGCGNLFPPNOLPGMDMIKQEGD 1458
Db 1349 LQMPGMNTVCSQMDNDPALRHTGLYCNQLSTDLTKTDAD 1388

RESULT 12
ID 061202 PRELIMINARY; PRT; 1405 AA.
AC 061202;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE STERIOD RECEPTOR COACTIVATOR 1A.
GN NCOAL OR SRCIA.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;

```

```

RN [1]
RP SEQUENCE FROM N.A.
RA Kamei Y., Xu L., Heinzl T., Torchia J., Kurokawa R., Gloss B.,
  RA Lin S.C., Heyman R.A., Rose D.W., Glass C.K., Rosenfeld M.G.;
  RT "A cAMP integrator complex mediates transcriptional activation and AP-1
  RT inhibition by nuclear receptors.";
  RL Cell 0:0-0(1996).
  DR EMBL; U56920; AAB01228.1; -.
  DR MGI; MGI:1276523; Ncoal.
  DR InterPro; IPR001092; HLH_dlm.
  DR InterPro; IPR000014; PAS.
  DR Pfam; PF00989; PAS; 1.
  DR SMART; SM00353; HLH; 1.
  DR SMART; SM00091; PAS; 1.
  KW Receptor.
  SQ SEQUENCE 1405 AA; 152572 MW; BD50F8FD62D852A CRC64;

Query Match 27.6%; Score 2108; DB 11; Length 1405;
Best Local Similarity 35.7%; Pred. No. 2,9e-120;
Matches 572; Conservative 236; Mismatches 434; Indels 362; Gaps 65;

Oy 1 MSGGENTSPSPRAETFRKRECP--DOLGSPKRTTEKRNREOEKTYEELAEIIFANFD 59
Db 1 MSGLDSSDPANDSHKRGSPCDTLASS---TEKRREOEKTYEGLAEELSANISD 56
Oy 60 IDNFEKPDCAILKETVKQIRQIEOEKAAANIDEVOKSDVSTSGQYIDKDALGPM 119
Db 57 IDLSLTVKPDCKILTKIVDQIQLMKRMGEKSTDDVOGSDSISSSQVIEKESLPL 116
Oy 120 LEALDGEFFVNLGNAVSENAVTOYLRYNOELANKSVYSLHVGDTHEFKNLPLKS 179
Db 117 LEALDGEFFVNLGNAVSENAVTOYLRYNOELANKSVYSLHVGDTHEFKNLPLKS 176
Oy 180 IVNGSGSGEPFRNSHTFCRMLVKPLPDEEGHDNOEAHQKTYETMOCAFVSOPEK 239
Db 177 LVNVPVPOEATRNSTHTFCRMLHP--PEDGTENOEACQYEVMOCFVYSQPSIQ 233
Oy 240 EEEGDLQSLICVARYRPMKERPVLPSESEFTTRQDLOGKITSLDITSTRAAMPQWEDL 299
Db 234 EDEGDFQSCILCARLP--RPAITGVESFMTQDTGKIISIDTSSLAAGTGWEDL 291
Oy 300 VRCTIQFHAHQEGESVYAKRHHEVLRQGLAFSQIYRSLSDGTLYAAQTSKLIRSQ 359
Db 292 VRCTIYAF--QPGREPSYAROLFQEVMTGTGLASSPSYRFLNDGTMLSAHTCKKLCY 350
Oy 360 TTNEPOLVISLHMLHREONCVNMP--DLTGOTMKP--LNPSSN---SPAQALCSGN-- 412
Db 351 SPDQPFIMGIHITDRHS--GLSPQDDNSGMSIPRINP--SVNPGISPAHGVTRSSLP 407
Oy 413 PGQDMLTSSNIN---PPIGPKBQMGMP--MGRFGSGGGMNHV-----SGMQATT 457
Db 408 PSNNMTVSARVNRQSSDLNSSSHTNSNNQGNFGCSPGNQIYANVALNOGQAGSQTN 467
Oy 458 PQGSNTYALKMNSPQSSPGMNPQPTSMLSPRHMSGVAGSPRIIPSSQSP--AGSLHSP 516
Db 468 P-----SLTNLNSMEGTGIALSQ---FMSPRRANGGLTRARMSNNSPPNPITLSP 519
Oy 517 VGVCS--TGNSHYTNSLNAQALSEGHCVSLGSSLASPDILKMGMLQNSPVNMPNPL 574
Db 520 VGIITSGACNNNNRSTYINIPYISLOGANEGPNNSVGFAGSPFLRQMSQNSPSRLSQP- 578
Oy 575 SKMGLSDKDCFGLYGPSEGTGQAESSCHPGEOKET---NDPNLPVPAVSSERADQGS 630
Db 579 AKAESKSKELIASILNMIO--SDNSQNSANEKGPLDGLHNNDRL-----SEGD 628
Oy 631 RLHNSKQTKILLTLTKSD-QMEPSPLASLST-----NKDSTGSLPQSG---STHGT 681
Db 629 KY--SQTSHKLVOLLTTAEQOLRHADIDITSCKDYLSCTGTSSSASNSPAGTGPSSH-S 685
Oy 682 SLKEKHILRHLLDSSPDVLAKTAEATGKDLQSSSSAPASE---VTIKQEVVSP 737
Db 686 SLTERKHLRLOE-GSPSDITTLTSLVEPEKDSVPASTAVSOGSGASIKLELDA 744

```

OY	738	KKKE----	NALLRYLLDKDD-----	TKDIGLPETTPKLERLDSKTDPASTKLIAMKEKE	789
OY	745	KKKESKHQLRLYLLDD	DEKDLRSTPMILCDDVKVKEKE--OMPCENTNPTPMKPAPE	803	
OY	790	EMSPFGQDQSEELDN	LEILLDLON--SOLPOLF--PDTRGAPAGVNDQAIINLMOUT	847	
OY	804	EYKLESOSQFTADIDQ	FDQLLPTLEKAQPLSCETDRMDGAVTG-----	VSIK 852	
OY	848	AENSPVTPVGAOKTAL	RISOSTFNNPPGOLGRLHPNOLPLDITLQSPFGAPPEPIRN	907	
DB	853	AE-----VLPA	LQPTTANA-----PR-----LSNL-----	PELEL 879	
OY	908	SSPSTVLPQPCMGNO	GICNGLNGSITGMIGNSASRPTMPSGEMAFQSAVRVCA	967	
DB	968	TTSAMNRPVQCGMIR	NPAASIPMRSSQPGQROTLOSVMYNTGSPLENNMGPPQYSQQ	10277	
OY	980	EALINDQFQPG-----	AGDQ-----IP-----MANNT-----	L 902	
DB	993	TTINONKP-----	EDDCISSQL-----	919	
OY	1028	APPNOTAPWPESILP	IDQASFASONRQPFSSPDDLLCHPAAESPSDEGALLDLQ--YL	10855	
DB	920	-----	DELCCPTVEGRNDEKALBELVSYFL	946	
OY	1086	ALNRPDLEELDRALG	IPELYSQSAVD--PQFSSQSDSN--TMEQKAAPVFPQYASQA	11411	
DB	947	SGKETELAELEDRAL	GIDKILV--QGGGIDLVSEFRPQQAOTPLIMEDRPTLVSQPS	10050	
OY	1142	QMAQGSYPMQDPN	FRHTMGORPSVATLRMO-----PRGRLPRTGYLON--QPNOL	11899	
DB	1006	PFA--GLSGPFGQ---	WVRKPSGLGAMPVQVYVTPPRKTFSPNMGQDROTILNRPPAPNOL	10660	
OY	1190	RLQLQHLRLQAO---	ONRPLMNQISNVSNVNLTLRPG-----VPTQAPINQMLAQROR	12400	
DB	1061	RLQLQORLQCGQOOL	HMQRORAIINQFNAANAPVGMNRSRSGMOQITPPPLNQAMLAQOR	11200	
OY	1241	EILNQHILRQMGHQO	QVOQRTLTMNGQ--GLNMTSMYAPBSGMATSNRPIQANQO	12988	
DB	1121	ELYSQQRHQRKOITQ	---QOKPMLMHRQSGFNIPPS-----SGLPVQMGDPRLLQGAPOQ	11722	
OY	1299	FPFPPNPG-----	-----	13066	
DB	1173	FPYPNPTGNTNCP	PPRPASTPFSQDLAANPEASLATRBSMNRRMAGNMGQCGAGISPMO	12322	
OY	1307	--ISQPPDPC-----	FTGATTPQSPILSPMAHTQSPMNQSQANPAYQAPSDINGM	13566	
DB	1233	QNVFQYQGPGLV	POGEATEFAPSLSPSSSMVPMVPVPPQSSILQCPPTSGVQSP--DMKAM	12911	
OY	1357	AQGMGNGMSMPSQ--	QSPPHFGQANTSMYSNNMNTINYSMAITNTGMSMSMNOGTQISMT	14144	
DB	1292	QQGTGNNVNSFOAV	QSPQ---APAPQGVY--NNMSTVYSMAGNANIONNPMNGOMQOM--	13466	
OY	1415	SVTSVSTGSLSMG	PEQVNDPALRGNLPLPNOLPGKMDMKOEGD	14588	
DB	1347	--SSLQMRGMNTV	CSQMDNPALRHGILGXCNQLSSDILKTTAD	13888	
RESULT	13				
ID	P70365	PRELIMINARY;	PRT;	1447	AA.
AC	P70365;				
DT	01-FEB-1997 (Tremblrel. 02, Created)				
DT	01-FEB-1997 (Tremblrel. 02, Last sequence update)				
DT	01-DEC-2001 (Tremblrel. 19, Last annotation update)				
DE	NCLEARECEPTOR CO-ACTIVATOR.				
GN	NCOA1 OR MNRC-1.				
OS	Mus musculus (Mouse).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.				
OK	NCBI_TaxId=10090;				
RP	SEQUENCE FROM N.A.				
RP	11				

RA Zhu Y., Qi C., Calandra C., Rao M.S., Reddy J.K.,  
RT "Mouse nuclear receptor co-activator, MNRC-1,"  
RL Submitted (Jul-1996) to the EMBL/GenBank/DBJ databases  
DR EMBL; U64606; AAB06177.1; -  
DR MGI; MGI:1276523; Nc0ai.  
DR InterPro: IPR001092; HLH\_dlm.  
DR InterPro: IPR000104; PAS.  
DR Pfam: PF000353; PAS; 1.  
DR SMART; SM00353; HLH; 1.  
DR SMART; SM00091; PAS; 1.  
DR  
SQ RECEPTOR, 1447 AA; 156897 MW; 832E026426E9BA62F CRC64; 4

Query Match	27.6%;	Score 2106;	DB 11;	Length 1447;
Best Local Similarity	35.9%;	Pred. No. 4e-120;		
Matches 574;	Conservative 234;	Mismatches 438;	Indels 354;	Gaps 65;

QY	1	MSMOENENTSDSRATERRKRECP--DOLGSPSKRNTERKRNROENKYIIBELALLIFANFND	59
Db	1	MSGJGDSSSDANPDSHRSKPCDPTLASS---TEKRRROENKYIELEALLIASINSD	56
QY	60	IDNFNFKPDOKALIKETKQJROIKEDEKAAANIDEVOKSDVSTGGVGDICALPBM	119
Db	57	IDLSLVKRDCKILKKTVDQJOLMKRNQEKSTDDVQKSDISSSGVIEKESLGPL	116
QY	120	LEALDGEFFVNLBGNVVSSENVTOYLRYNOBELMKNVSYLHVGDHTEFFVKNLPS	179
Db	117	LEALDGEFFVNLBGRITVFESENVTSYLGYNQOELMTNTSVYSLHVGDHAFVKNLPS	178
QY	180	IVNGSGMSGEPRNRSHNFNCNMYKPLPDEEEDHGNQOENKUTMOCNPAVOPKIX	233
Db	177	LVNGVPVQOETRRNRSHNFNCNMLHP--PEDGTEHNOEQCNREVMQRTVSOPISTQ	233
QY	240	EDEGDELJCLICVARNVPMKERPVLPSESEFTTRDLOGKITSLDSTIMRAAMSGEDL	299
Db	234	GDGEDOFCCLICIARRBP--RPRAITGVESEPMKODPTGKITSLDSTILRAAGTGMEDL	299
QY	300	VRCIOQKHAOHEGESVYAKRHHHEVULROGLAFSOIYRSLSDGTYLVAOATKSLIRSQ	355
Db	292	VRCICTYAEF--OPROGEPYSARQLOEOWTRCTASSPSYRFLINGTMTLSAHTKCKILCPQ	350
QY	360	TTNEPOLYISLHMLHREBNCVMNP--DLTGQTMKP--LNPISN---SPAHQALCSGN--	412
Db	351	SPDMQPFMTGHIHIDREHS--GLSPDODDSNGSMKIPRIN--SVNPGISPAHGTVRSSTLP	407
QY	413	PGODMTLSSNNI---FPINGPKQOMGNP--MGRFGSGGGMHNVSGM---QATTPQGS	461
Db	408	PSNNNVYASARNRQOSSDLNSSSHHTMSSNNQMGFGCSPGMOIYANVALNQGQAGS--QSS	466
QY	462	NYALKMNSPQOSSPFGMANPQOFTSMKSPRHMSPGVAGSPRIIPSOFSF--AGSLHSPVGYC	520
Db	467	NPSLNLNNSPMEGCTIALSQ---PMSPRROANSGLAFRRARNSNNSFPENPIPTLSSPGIT	523
QY	521	SS--TGNSHSYTNSLNAOLALSGHGVSLGSSLASDMLKGNLONSPPVNNPPRLKMG	578
Db	524	SGACNNNNRSTSNIPVTSLOGMANGPNNNSVGFSGSPVLRQMSQNSPRLSMOP--AKAE	582
QY	579	SLDSKDCFGLYGEPEEGTGOAESCHPRGEOKET---NDPNLPRPAVSSERADQOSRLHD	634
Db	583	SKDSKEIASILNEMIQ--SDNSDMSANBGKPLDGLLHNDRL-----SEGDSKY--	630
QY	635	SKGQTKLLOLITTSF--OMEPSPLASSLSDT-----NKDSTGSLPGSG--STHGSLAKE	685
Db	631	SQTHSHKIVOLLTTTAEODLRHADIDTCKDVLSCGTGSSSASNSPBGCTPPSSH--SSLTE	669
QY	686	KHKILHRLQDSSPVDLAKLTAAFGTKDLOESSSTAPAGE---YTIQOEPVSPKKE	741
Db	690	RKHLIHLRLPR--GSPSDITTLTVSPEPKKDVSPASTANVSYGQSGKASLILELDAKAKE	748
QY	742	---NALLRYLLDKD---TKDIGLPEITPKLERLDSKTDTPASNTKLIAMKTEKESMF	793
Db	749	SKDQOLRYLLDKREKDLRSPNLCIDVDYKVKYKKE--QMDPCNTNPTPMTPKRAPPEVYL	807

QY	794	EPGDQSGSELDNLEELIDLDLQ--SQLPQLF--PSTRGAPAGSVKQALIIINDLMQLTAE	NS	851
Db	808	ESQSGFTADIDPOFPOLLPTLEKAAPSLSCENTRMDGAATG-----VSIAKAE--		854
QY	852	PYTPRGAKTALRTLSQSTFNPNRPQGLRFLRNQNLPLDITLQSPRGAPPPPIINS	SP	911
-Db	855	-VLPAISLQPTTARA-----PR---LSRL-----PELELEID		883
QY	912	SVTPDQGMNMGQMGIMNOGMLGNSSTGCMGNSASRPTMPSGEMADQSSAVRVTCA	TS	971
Db	884	NOFGQPG-----AGDQ-----IP-----WANN-----LTTIN		906
QY	972	MNRPVQGMITNPAASLTPMRSSQPGQROTLDQVWNNITGPSELENNMGPGYSQQA	AP	1031
Db	907	QNKP-----EDQISISQL-----		919
QY	1032	QTAPRESILPILDQASPASQNRQPGSSPDLLCPRPAESPSDEGALLDOL--Y	LA	1089
Db	920	-----DELCCPTTYEGNDEKALLLEQLVLSLSKD		950
QY	1090	FDGLEEDRALGRLDELVSQSOADV--PEQSSQDSN--IMLEQKAPVRPQOYASQAQ	MAQ	1145
Db	951	ETELAEIDRALGRLDKLV-QGGGLDVLSERPQOAPRPLMMDRPTVYSQPYSSPS	PA	1008
QY	1146	GSYSPMODRNTHTGQPSYATLMO-----PRGLRPTGLVQN--QPNQLRLQ		1193
Db	1009	GLSGPFG--MYRQKPSLGAMVUOTVPRGTFSNMKQMPQOTLNRPAPRNQLRLQ		1064
QY	1194	QHRLOAQ-----QNRQPLMNOISVNSVNTLTPRG-----VPTQAPLNAQMLAQR	REL	1244
Db	1065	QQRLOQGGQQLHQRKQALINQFAPANPVGMNMSMGQQLTPQRLNMLMLAQRREL	YS	1124
QY	1245	QHLRQMHQOOQYQORTLMMRQ--GLNKTPSMAVSCMPATMSNPRIQANAOQPP		1302
Db	1125	QOHRQRLIQ--QQRAMLMRQSFQNNITPS--SGLPVQMGTPSLPQAPQPPYR		1176
QY	1303	PNVY-----IS		1308
Db	1177	PNYGTNPCTPASTSPFSQLANPEASLATRSSMVMKRGAMGNGQFGAGISPMQONV		1236
QY	1309	QQPDPG-----FTGATTPQSPILMSPRMAHTQSPDMQSOQANPAYQAPSDIN	GAQ	1360
Db	1237	QYRPGGLVPRQGEATFAPSLPSSGMVPMRPPQSSLLQOTPPTSYGQSP--DM	KAMQ	1295
QY	1361	MGGSMSFQ--QSPPHFGQOANTSMNSNNINVSATVTGKSSMNQWTQISMTSV		1418
Db	1236	MGNNVNFSQAVQSP--APAQPPV--NNKSTIVSAGAGNANIQNMNPMQOMQ--	SS	1348
QY	1419	VSTGLSMGEGVNDPALNGNLFNPQMGMMIOEGD		1458
Db	1349	LQMRMNTVCSEQNMDALNHTGLXCNQLSSTDLTKTDAD		1388
RESULT 14				
Q9EPU2				
AC	Q9EPU2:	PRELIMINARY;	PRT:	1082 AA.
DT	01-MAR-2001 (TREMBLrel. 16, Created)			
DT	01-MAR-2001 (TREMBLrel. 16, last sequence update)			
DT	01-DEC-2001 (TREMBLrel. 19, last annotation update)			
DE	NDCLREAR RECEPTOR COACTIVATOR AIB1 (FRAGMENT).			
OS	Rattus norvegicus (Rat).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
OX	NCBI_Taxid=10116;			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=SPRAGUE-DAWLEY;			
RA	Bigsby R.M., Long X., Nephew K.P.;			
RT	"p98059 induces estrogen receptor-coactivator interaction and			
RT	reported gene transcription."			
RL	Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.			

DR	EMBL, AF322224; AAG42837.1; -,
KM	Receptor.
FT	NON_TER 1 1
SQ	SEQUENCE 1082 AA; 115131 MW; BA35D946C8C8C080 CRC64

Query Match	22.1%;	Score 1684;	DB 11;	Length 1082;
Best Local Similarity	37.1%;	Pred. No. 1.7e-94;		
Matches 448;	Conservative 177;	Mismatches 381;	Indels 200;	Gaps 44

OY	239	LYRRCLQFPAHQEGESVSTAKRHHEBVLROGLASQITRFSLSOGTLVAQTKLIRS	358
OY	239	LYRRCLQFPAHQEGESVSTAKRHHEBVLROGLASQITRFSLSOGTLVAQTKLIRS	358
Db	1	IIRRCIOFRFSLNDQSGWS-QKRHYOEAYTHGAETPRVFRSLADGATIVASQTKSPFRN	59
OY	359	QOTNEPOLVATLSLHLEHQVQVYMNPDLTNGQTMGKPLNTSSNSPNAHQASGNGOQMT	418
Db	60	PVYDNRHGEVSTHFLQREQNCRRPNTL-----ODKIRP-----PAAGGMSLSVSQVQ	110
OY	419	LSNINFEPPINGPKQOMMPMGKRFEGSGGGMNHYV-GMOATTP---OGSNATLKMNSPOSS	474
Db	111	MLGSRITGVADPSTGTGMAGARYGASSVASLTPGQSLQSPSSYQNSYGLMMSSPPIHGS	170
OY	475	PGMNPQOPTSKLSRPHMMSGVAGSPRIPIPSQFSPAGSIHSPVGCSSGTGNSHTTNSL	534
Db	171	PGLPNOONTIMISRRNR-----GSPRMAHQSPAGVASHPMGSSGNTG--SHSPSSSL	223
OY	535	NALQALSEGHVSLGSSLASPDLMKQNLQSP--YNNNPPLKMSGLSKDQFGLYGPSS	593
Db	224	SALQALISEGVGTSLTSLSSPGKR---LNDSPMNNTNQP--SKASSQDSKSPGLGICQON	278
OY	594	EGTTGQAESCHPEQKEETDPMLPAPVSSERADQOSRLH-----DSKQGTCLLQTLT	646
Db	279	-----PVESSYCPNSMD-----PVYTKENKENGSEASETPRGPLESKHGKLLQTLT	326
OY	647	TKSQMEPSPLASLSTN-KDS-----TSLSPGSGSTHGTSLKEKHKLH	691
Db	327	CSSDQRHGSJLNSPLDSONCKDSIVTSPSGVSSSTSAVSTSMHGSLSLOEKHRLH	386
OY	692	RLLDSSSPVDLAKLTAEATGKDLQDSSSTAGSEVTIKOEVPSPKKE-NALLRYLLD	750
Db	387	KLLONGSPAEVAKITREATGKD-----TSSSTAGSGGSYXOELSPXKKENALLRYLLD	442
OY	751	KDTRKDGLPEPTPKLERLDSK-----TDPASNTKLTKMTKEKMSFEPEGDPGSEL	803
Db	443	RDPSPVLAKELQADQAGDSKLSQSCSXNPSSQGEK-DPKITHEVVS-----GDL	493
OY	804	DNLEELLDDQNSOLPOLFPDTPRGAPAGSVDKOAILINDLMQITANSEVYT---PYGAQK	860
Db	494	DNLDAITGLDITSSDPF-----YNSPTNGSHHGAAQO	523
OY	861	TALRISOSTENNPPQGLRLPNOPLDITLQSPGAGPPPIRNSPSEVIR-OPGM	919
Db	524	MFAPGSSSLGLRSPQVOSVTPPYNRALSLD---SPVSYGVSPPYKANSAPFVLPKQPTL	579
OY	920	MGNQMGINGNGLNNSGTGMIGNSASHPTMPGSEMA-POSSAVRYTCAAT-----	968
Db	580	AGNFRMMDQENGCANNGGNRRVNPVNPSTSSQDGLNARSASRMEPLASSPLGRAGDY	639
OY	969	TSAMNRPVQGMATRNPASTIPMRPSSOPGQROTLAGSOVANNIGPSELKMMGPQYSSQQA	1028
Db	640	SAALPRPALG---SSGPTPLPLRSNRLPGARPTMLQ-MRAG-EVPMGMGVSPYS-PAY	691
OY	1029	PNQNTAPMPESTLPIDQASASQNRQPFQSSPDLLCIPHAESPDSDEBALLDOXLXR	1088
Db	692	PSNPGSNPBGMSMEGPHGAONRPLRLKRSDELGLGPPSNBEGQSDERALLDOHTLLS	751
OY	1089	NFD---GLEEIDRALGIDELVSOQASQVADPE--FSSQDSNIMEEOKAPVPQOQYASQQA	1144
Db	752	NTDATGTEIDRALGIDELVSOQALESQKQDYVQGEAVALMMDQKRALYGCOTYPRAGQPL	811
OY	1145	QGSYSPMODPNHTMGORPSYATLRMQ-----PRRGLRPTGLV---QNOPDLRLQL	1193
Db	812	QG-----GFHIOGSPSLNMSQISQOGSFPLOGJHPRASHVNRPTTNPQKLRMQL	863



OC Mammalia; Eutheria; Primates; Catarrhini; Homiinae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Raineri I., Soler M., Senn H.,  
 RT "Analysis of Human Immunodeficiency Virus type 1 promoter insertion in  
 RT vivo."  
 RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; U19179; AAA64187.1; -  
 SO SEQUENCE 416 AA; 45239 MW; 68AF955508339CAD CRC64;

Query Match 7.1%; Score 545; DB 4; Length 416;  
 Best Local Similarity 33.6%; Pred. No. 2.1e-25;  
 Matches 142; Conservative 63; Mismatches 110; Indels 108; Gaps 16;

QY 1126 LEKAPVFPQGVASQAQMGSGSPMDPFHMGQRPSTATLRMPRP-----GL 1176  
 DB 1 MEERPNLYSQPSSPSPTAN-----LPSPQGVNRQKPSLGTMPVOYTPRGAFSPQMG 55  
 QY 1177 RPTGLVON--OPNQLRLQHLRLAQ-----ONRQPLMQISNVSNVTLRPG-----V 1224  
 DB 56 QPQOTLNRPAAPNQLRLQHLRLAQGLIHRQALILNQFATAPVGINMMSGMOQOI 115  
 QY 1225 PTOAPINQMLAQORERILNQLRQRMHQOQVQOFTLMRGO--GLNTPSPMVAFSGM 1282  
 DB 116 TPQPLNAQMLAQORERILNQLRQRMHQOQVQOFTLMRGO--GLNTPSPMVAFSGM 167  
 QY 1283 PAMNSNRITQANAAQGFPPFPNGISQPPRPTG-----ATTPQSLT----- 1325  
 DB 168 PVQMGNRLPQAGAPQOQPPYPNGTNGTPPASTSPESQLAANEASLANRSMVSRGMT 227  
 QY 1326 -----MSPRMAHT-----QSPMAQ 1340  
 DB 228 GNIGGFGCTGINTNQMOQNVQYPGAGVPGGEANFAPSLSPSSMVMPIRPPQSSILQ 287  
 QY 1341 SQANPAYQAPSDINGMAQMGNSMFSQ--QSPPHFGQOANTSMSTNNINVSATNT 1398  
 DB 288 TPASGQSP--DMKAWQQAIGNNVFSQAVQNP--TPAQGVY--NNMSTVSMAGN 342  
 QY 1399 GGSMSNMQMTGISMSTSVSTSGLSMGEQVNDPALGSLNLPQDLGMMIKQEGD 1458  
 DB 343 TIVQNNPMAQOMQOM--SSLQMPGAMTVCPQIINDPALNHTGLYCNHLSITDLKTEAD 399  
 QY 1459 TTR 1461  
 DB 400 GTQ 402

RESULT 17  
 Q9BR49 PRELIMINARY; PRT; 203 AA.  
 AC Q9BR49;  
 DT 01-JUN-2001 (Tremblrel. 17, Created)  
 DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)  
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
 DE D1049G16.2.2 (CONTINUED FROM BA456N23.2 IN EM:AL353777 AND D1237J2.1  
 DE IN EM:AL021394) (FRAGMENT).  
 GN NCOA3.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiinae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Skuce C.;  
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AL034418; CAC36067.1; -  
 DR InterPro; IPR000014; PAS.  
 DR SMART; SM00091; PAS; 1.  
 FT NON\_TER 1  
 SO SEQUENCE 203 AA; 23303 MW; 895D62D858C3B04C CRC64;

Query Match 6.7%; Score 508; DB 4; Length 203;  
 Best Local Similarity 56.7%; Pred. No. 1.4e-23;  
 Matches 102; Conservative 24; Mismatches 44; Indels 10; Gaps 2;

QY 122 ALDGEFFVNLGCVVSEENVQYLRYNQDELANKSVSILHVGHTFEVKNLPSKIV 181  
 DB 1 ALDGEFFVNLGCVVSEENVQYLRYNQDELANKSVSILHVGHTFEVKNLPSKIV 59  
 QY 182 NGGWSGEPFRNRSHFENCMLYKPLPDSGEEGHQDEAHOKETQCFVAVSPKSKKEE 241  
 DB 60 NGVSWTNETQKSHFTFNCMLKTPHDLIEDINASPEKQRETQCTALSPRAMME 119  
 QY 242 GEDLQSLCICVARNVMEKRPVLPSSSEFTTRQDQKITSLDTSTMAAMPQWEDLVR 301  
 DB 120 GEDLQSCMLCVARITTGERTFPSPNPFITRLDSDV-----LLCCGSGAMVR 170

RESULT 18  
 Q9GS19 PRELIMINARY; PRT; 2035 AA.  
 AC Q9GS19;  
 ID Q9GS19;  
 DT 01-MAR-2001 (Tremblrel. 16, Created)  
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)  
 DE BCDYSONE RECEPTOR CO-ACTIVATOR TAIWAN.  
 GN TAI OR CG13109 OR CG18494.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Bai J., Uehara Y., Montell D.J.;  
 RT "Regulation of Cachein-mediated Cell Motility by Taiman, a Drosophila  
 RT Protein Related to Atgbl, a Steroid Receptor Coactivator Amplified in  
 RT Breast Cancer."  
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AY008258; AAG1637.1; -  
 DR FlyBase; FBgn0041092; tai.  
 DR InterPro; IPR001092; HLH\_dlm.  
 DR InterPro; IPR00014; PAS.  
 DR SMART; SM00353; HLH; 1.  
 DR SMART; SM00091; PAS; 2.  
 DR PROSITE; PS00583; PKB\_KINASES\_1; UNKNOWN\_1.  
 KW Receptor.  
 SO SEQUENCE 2035 AA; 213790 MW; 6EC1F9E8447B1A5 CRC64;

Query Match 6.5%; Score 495.5; DB 5; Length 2035;  
 Best Local Similarity 19.9%; Pred. No. 2.2e-21;  
 Matches 365; Conservative 218; Mismatches 623; Indels 627; Gaps 79;

QY 1 MSGMGENTSQPSAETRRKRECPDQGPSP--KRNTKRNREQENKYEELAEIAPNPN 58  
 DB 209 ISANSANSKATSGR-KIRBRKDSKVNLPOQIKNCKNEKKRREAENGVIYQLSILFLNKR 267  
 QY 59 DIINFNPKPKCAILKETVQIQI--KEQEKAAA-- 92  
 DB 268 G-DMTSTPKDKAAILQVYVTFREIDKQGNRDISSTNNNSTTTNNNTNSNNNNT 326  
 QY 93 -----NIDVQKSDVSTGGGVIDKD-ALGPM-----MLELDGF-- 126  
 DB 327 SKQOATSTKSCRAITDNCSTHPVQGEVSTPEPDSILLQVETISYFALPHYIS 386  
 QY 127 -----FFVNLGCVVSEENVQYLRYNQDELANKSVSILHVGHTFEVKNLPSKIV 181  
 DB 387 GVGWVLLQVWANGIIESCQINRDLIGYERQELYHQPLVWYILVSGHAKL-----EPIT 440  
 QY 182 N-----GGSNS 187

D	b	441	NTMNNP	NGNS	SNNS	ANNS	GGSS	AGT	SAC	YAC	W	D	E	L	N	G	N	A	S	O	G	S	N	S	G	A	G	G	500																										
Q	y	188	GEPR	NR	S	H	T	F	N	C	M	L	V	K	P	L	P	D	S	E	-----	E	E	C	H	D	N	E	A	H	O	-----	K	Y	E	T	M	O	C	F	A	V	S	O	P	K	S	I	K	E	240				
D	b	501	AAAK	K	R	S	I	S	T	K	V	M	L	K	D	T	R	T	A	T	O	T	T	S	N	C	E	E	K	P	L	R	O	S	H	O	K	Y	E	E	V	L	I	A	-----	P	K	D	556						
Q	y	241	E	G	E	D	L	O	S	C	L	I	C	V	A	R	R	P	M	K	R	P	V	-----	L	P	S	E	S	T	T	O	D	L	O	G	K	T	S	L	D	I	S	T	M	R	A	291							
D	b	557	DA	-	D	A	S	S	V	L	C	L	I	T	R	-	P	E	D	S	P	L	E	I	N	I	O	H	V	O	O	P	I	E	O	T	F	K	I	D	I	G	K	L	T	L	D	P	A	L	R	P	614		
Q	y	292	M	K	P	E	M	E	L	V	R	C	I	O	K	F	A	H	O	E	G	S	E	S	V	A	K	R	H	H	E	V	L	R	O	L	A	-----	F	S	O	I	R	F	S	340									
D	b	615	F	K	H	L	O	T	W	B	R	L	M	O	D	L	C	H	P	D	-----	L	S	T	L	S	H	L	E	D	I	O	D	S	A	N	S	P	A	G	A	C	T	S	V	A	S	P	R	I	R	671			
Q	y	341	L	-	S	D	G	T	L	V	A	O	T	S	K	L	I	R	S	O	T	T	N	E	P	O	L	V	I	S	H	L	H	R	E	O	N	-----	V	C	V	381													
D	b	672	L	G	A	D	V	Y	V	H	K	A	N	S	R	L	F	L	N	O	T	P	E	G	O	F	I	N	S	V	O	L	I	N	S	E	N	D	M	S	S	M	T	G	A	S	G	L	G	I	O	L	C	A	731
Q	y	382	M	N	E	D	L	T	-----	G	O	T	M	K	L	N	P	I	S	S	-	P	A	H	-	O	A	L	C	S	G	N	-	G	O	M	T	L	419																
D	b	732	M	A	P	S	P	L	A	S	L	S	L	S	M	D	G	L	H	G	T	G	S	S	S	P	A	S	A	S	M	L	P	T	H	L	L	G	I	V	G	G	O	O	G	G	N	S	T	O	791				
Q	y	420	S	S	N	I	N	F	-----	I	N	G	P	-	K	R	O	M	P	M	G	R	G	S	G	S	G	M	H	-----	V	S	G	M	A	T	-----	456																	
D	b	792	T	T	S	V	G	L	M	S	A	I	I	N	G	L	O	O	O	O	O	O	O	O	R	G	S	A	S	S	A	S	A	N	A	L	V	M	A	F	T	A	S	P	A	P	A	E	H	S	F	Y	G	S	851
Q	y	457	-----	T	P	O	G	S	N	Y	A	L	K	-----	M	N	S	P	O	S	S	P	G	M	N	P	G	O	P	T	S	M	L	S	P	R	-----	H	R	M	P	G	A	S	P	R	501								
D	b	852	D	T	R	E	F	D	L	A	M	I	S	S	F	E	L	D	P	S	G	V	A	M	T	D	S	R	P	N	S	A	S	A	T	V	T	S	T	P	R	P	S	G	H	S	P	A	V	C	A	S	P	A	911
Q	y	502	I	P	S	O	F	-	P	A	G	L	S	H	P	V	G	C	S	T	G	N	S	H	S	T	N	S	L	A	L	A	L	S	E	G	H	V	S	L	A	S	P	L	M	K	G	560							
D	b																																																						

Qy	1110	QAVDEQSSSDSNIMLEQKAPVTPQGYASQAQMAQGSVSPMDPFTHTMGCRPSYATLR	1165
Db	1369	SAINDIQ-----KSLIMLDVESAAFGNDLNOILMMTQ-----QOQHOOQOQOQOQLLALQL	1417
Qy	1170	MQPRGLPPTGLVQ--NOPMDLRL-QLQHHLQAQORPLANOISVSN-----	1215
Db	1418	AQOQOQOQROQHILQOPRAYPGMLNMQOQOQHOOQOQOQOQHITMORLEMRKQMGQGRPP	1477
Qy	1216	-----VNLTLRPGVTPQAPRINAMQLAQOREITLNOHLRQMHQOQOQVQR-----	1261
Db	1478	MYPARGREPMNAVATPG-GVYLPAQOQLRNIROQOQLAAQQKEXRLILQOQOQLLVEN	1536
Qy	1262	-----TLMMQO--GLMTPSVNAVPSGMP-ATMSNPRIQANAAQCP	1300
Db	1537	ASEYWGNSNLIQLIKLIVSYILAGMNAAGLNINISLTLNTTGAPVNSLSRTNLP-SDAQ-	1592
Qy	1301	FPFNVG---ISQODPGEFTGATPQO-----SPLMSPRAAHTQSPMOOS	1341
Db	1593	LSPNFAQTLMOQQLSPGRSAPYSQPQPGYAPQFPQPGQRLSPFOOQOQLSQOQNNVQOQ	1655
Qy	1342	Q-----ANPAYQAPSDING-----WAQGNMGNS-----MF	1367
Db	1653	QLAVQOQOQVGDGSGSRNTPFGSNSQMSQPMQNSPQMGSGGGGGGPGPLPSGNAGRLL	1712
Qy	1368	SQSPSPHFGQQAQNTSM-----SNMNTNVSMAATNTG-----MSNMQNTGQISM	1413
Db	1713	QOHNMLTALQAGVSPYNARQOONORRGINSPGAAGPGNPAQAALORONSFGOG--GG	1771
Qy	1414	TSVTSVSTSGLSMGPEQVNDPALRGNGLEPNQ	1446
Db	1772	GGATTPEGSGVGFQGPQ---SPGTGVNVNFQOQ	1801
RESULT	19		
Qy	QVYLD8		
AC	QYVLD8:	PRELIMINARY;	PRT: 1778 AA.
DT	01-MAY-2000 (TREMBLrel. 13, Created)		
DT	01-MAY-2000 (TREMBLrel. 13, Last sequence update)		
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)		
DE	CG18494 PROTEIN.		
GN	TAI OR CG13109 OR CG18494.		
OS	Drosophila melanogaster (Fruit fly).		
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;		
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachyoptera; Muscomorpha;		
OC	Ephydroidea; Drosophilidae; Drosophila.		
OX	NCBI_Taxid=7227;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=BERKELEY.		
RX	MEDLINE=20196006; PubMed=10731132;		
RA	Adams M.C., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,		
RA	Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,		
RA	Georg R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,		
RA	Sturgeon G.G., Wortman J.R., Randall M.D., Zhang O., Chen L.X.,		
RA	Brandon R.C., Rogers Y.-H.C., Blazer R.G., Chame M., Pfeiffer B.D.,		
RA	Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.K., Miklos G.L.G.,		
RA	Abdil J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,		
RA	Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,		
RA	Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,		
RA	Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,		
RA	Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,		
RA	Cherry J.M., Caley S., Dahlke C., Davenport L.B., Davies P.,		
RA	de Pablos B., Delcher A., Deng Z., Duan A.D., Dew I., Dietz S.M.,		
RA	Dodson K., Doup L.E., Downes M., Mayas-Rocha S., Dunkov B.C., Dunn P.,		
RA	Dubin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,		
RA	Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,		
RA	Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,		
RA	Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,		
RA	Hoslin D., Houston K.A., Howland T.J., Wei M.-H., Ibergman C.,		
RA	Jaitani M., Kalush F., Kapen G.H., Ke Z., Kennison J.A., Ketchum K.A.,		
RA	Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,		











```

Db 687 GAAVDLLTDPDYPDPATQQLITKLVTGPMACGAVMPTTMMGCVTAFAPOGQOPOT 746
OY 1002 LOSOVNIGPSELEMMNGGPOYSQOQAPNOTAPWPESTLIPDQASFASONROPFGSSPD 1061
Db 747 I-----SIAQPPASQTA-----DQQTHTQAQTQ-----769
OY 1062 DLLCPHPAESPDEGALLDQLYLALNFDGEEIDRALGIPELVYSQSAVDEQFSSSD 1121
Db 770 -----AAATAQOQOQONQOL-----TQOQTQFLOAPRLL-----HSNQS 803
OY 1122 SNIMLEOKAPVFPQOYASQAQMAQGSYSPPMDNFHTMGRRPSYATLRLMPRRGLRPTGL 1181
Db 804 TQILITD---AAFPLO-----OQGFPTATQO-----826
OY 1182 VQNPQPLRLQDRLQALQAOQNPQPLMNQISNVSNVLTLPFGVPTQAPLINAQMLAQORE 1241
Db 827 -QOQLHQOQOQOQLOQOQOQLOQOQO-----0000Q 853
OY 1242 ILNQLRLRQRMHQOQOQOQFTLMMRGGLNMT 1273
Db 854 LQOQHQQOQOQOQLOQOQOQOQOQLAHRSDSMT 885

RESULT 24
O9GRA9 PRELIMINARY; PRT; 2310 AA.
AC O9GRA9;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE MASTER OF THICK VEINS.
GN SBB OR MTV OR CG5580.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxId=7227;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE-LARVAL DISCS;
RA Funkoshi Y., Minami M., Tabata T.;
RT "mtv, a novel gene that shapes the activity gradient of the DPP
  morphogen through regulation of chick veins."
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB044403; BAB20792.1; -.
DR FLYbase: FBgn0010575; sbb.
DR InterPro: IPR003880; Phosphopant_attach.
DR InterPro: IPR000822; Znf_C2H2.
DR SMART: SM00355; Znf_C2H2; 1.
DR PROSITE: PS00012; PHOSPHOPANTETHEINE; UNKNOWN_1.
DR PROSITE: PS00157; ZINC_FINGER_C2H2_2; 1.
KW DNA-binding; Zinc-finger.
SQ
SEQUENCE 2310 AA; 230954 MW; 320EAB65E51AEDF6 CRC64;

```

Query Match 4.4%; Score 339; DB 5; Length 2310;  
 Best Local Similarity 20.6%; Pred. No. 9.9e-12;  
 Matches 278; Conservative 150; Mismatches 499; Indels 422; Gaps 61;

```

OY 369 SLMLHREQVCMYNDPLTQGTGKPLNPTSSNSPA--HVALCSGNQGMITSSNINF- 425
Db 978 SLTFSHAQVSHKLRN-----GATKRGATRSASNAANSSSSSGNGGATPSTSPATFL 1033
OY 426 -----PINGPKEDQMHPMRFGSGGKNHYS-----GMQATTPOG 460
Db 1034 PPRPEKRSKDEAPSLNGASD-GASGGIGGAGGYMVMYNASGIPRLSASGGGLATDPQS 1092
OY 461 -----SNYALKMNSPSQSSPGM--NFGQPTSM-----LSPRHRMSPG 495
Db 1093 LLMPTVGLNVQISTKCKKTASPCALISPVLLCEPDQCSKTKYKHAHGLRYHQSHAHGAGGG 1152
OY 496 VAG-----SPRPPSQFSPAGSLHSPVGYCSSTGSHSYTNSLVALQALSLSGHG 545

```

```

Db 1153 ASSMDEDSQAPEDPATPPSP-----GVASGTSSGASVASSAVPA-TAPBAGQ- 1199
OY 546 VSLGSSLASPDLLKMGNIQNSPVNMNP-PLPSKMSLDSKDCFLGYPESEGTTGQAE--- 601
Db 1200 ---GTVAVSPNTPLANSSNPVTNVAPASPATGSVTIAAINTPTSVVETQAPRLGPPRVP 1256
OY 602 -----SSCHGEQKETND---PULPPAVSSERADGGSRLHDSGGQTKL---QLLT 646
Db 1257 PAPTPICAVATPGEASVSVLPLGLNLPITAGPNSATQOQPPQOQQLLVPGGSAAS 1316
OY 647 TKSDQMEPSPLA-----SSLDTNKDSTGSLPGSGSHTGTSLKKEKHLKHLHLD 696
Db 1317 LQOQOQOQOQOQVAGSGITAGISGQALSQHQOQLMGLP-----AMLSQOQ--QALLQ 1367
OY 697 SS-----SPVD---LAKTFAEATGKDLQESSTARGSEVYTIKQEP----- 734
Db 1368 GALKAGVLRGPRDGNPLQOQPGASVNPQTQOSPFRPP-SHQODQGTFSATVQAQAGLKT 1426
OY 735 -----VSPKKEN-----ALLRYLDDKDDTKDIGLET---TP-KLERLDS 771
Db 1427 SPFGSGVGAASSKOKKKNKSPGSPDFEGRVREDVQSPAYSDISDSTPVAEQEMLDK 1486
OY 772 KTDPASNTK---LIAMKTEKEMSFEGDQPGSELNLEIIDLQNS-----QLDQ 820
Db 1487 SVGAQVATKHTELMGKRPTEVGVPRAAPRNMYGKTOFTYPAQOQASAPRPOQOQOQO 1546
OY 821 LFPDTPGAPAGSVDRQALINDLMQLTAEHNSPYT-----PVGAOKTALRISOSTEN 871
Db 1547 YMYQTEPRGKPRGLPALQQAQOQOQLOQPGAPRPTSQPSPHLLGPPQOQVAHLADYSGK 1606
OY 872 N-----PRG-----QLGRLIPNQLPLDI-TLQSPGACGFPPIRMS--- 909
Db 1607 NKDPPLDLMTKPPQPPQPPSQOQSGQSLGQDENNGKDVGPPTSQPSQP-PRVNLISAVA 1665
OY 910 --PYSVIQPPG-----MGNQMGIG-----NOGNIGNS 936
Db 1666 GPPPGSLP-PGLGGLSALGALGAGCGEPGKGMHFTFPNFTIPATRYYNVNDPAGSVIYA 1724
OY 937 TGMIGNSASRPTPSPGEMAPQSSAVRYTCAATTSAANRPYQGMIRNPAASIPMRSSOP 996
Db 1725 SEBAKLSGHPRLPSPSQAOQLSGISIK-----EERLKPSP--SPHDQKHM 1770
OY 997 GQRTQLQSOVMNIGP-SELEMMNGGPOYSQOQAPRNPOTAPWPESTLIPDQASFASONROP 1055
Db 1771 SQOQMIASKLIKQEPMTKQIKQEPNSNPGQCHPRPQOQPARP--PQOQOQPPPPPOQ 1827
OY 1056 FGSSPDLCRHPAASPSDEGALLDQLYLALNPFGLIEDRALGIPELVYSQSAVDPE 1115
Db 1828 HALHPKDL-----OALGATYPAITYORHSL--- 1850
OY 1116 QFSSQDSNIMLEOKAPVFPQOY-----ASQAQMAQGSYSPPMDNFHTMGQORPS 1164
Db 1851 -----NLAVQQAAREBELRLRYMYFTGRQNSAAAAAANAS-----GGLRP 1893
OY 1165 YATLRMQPRFGLRPTGLVQNPQRLQL-QHRLQ-QQNRQPLMNQISNVSNVLTLP 1222
Db 1894 HPQMHKDEFGMGMSAAQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQO 1940
OY 1223 GVTQAPLINAQMLAORQREILNHLRQRMHQOQOQOQOQOQOQOQOQOQOQOQOQOQOQO 1273
Db 1941 -----HQAQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQO 1993
OY 1274 -----PSWAPSGMPATMSNPRIPOANAQOFPFPNYGISOQPD-GETGAT 1319
Db 1994 KDSPPKQGGDDQPLAKYKQGGQKPTMETQGRPPRPSPQYLYHSY-IS-PTPFGF--- 2046
OY 1320 TPQSPILSPMAHTQSPPMQOQANP---AYQAPSDINGAGQGNMGNSM-----FSQ 1370
Db 2047 DPNHMPYRNVLMASAGGYNTAPYHLRIPRYHAHREDIS-----NNTGKALDALHHAQY 2102
OY 1371 SPFH-----FGQANTSMYSNNMINIVSMAT-----NTGGMSSMNMQTOGQISMTSVTS 1420
Db 2103 YTHKIHETLSERALKSPTSKSGPVKVSVPISGPPQGGTSSGPGSGHVS----- 2154

```

```

Oy 1421 TSGLSMGPEQVNDPALRGNLFPNQLPG 1449
Db 2155 -----GVLAGPGS-----GSMQPGOSAPG 2172

RESULT 25
O9NJ17
ID O9NJ17 PRELIMINARY; PRT: 5476 AA.
AC O9NJ17;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE SPLIT ENDS.
GN SPEN OR CG18497.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydrioidea; Drosophilidae; Drosophila.
OX NCBI_Taxid=7227;
RN [1]
RP SEQUENCE FROM N. A.
RX MEDLINE=20157049; PubMed=10655223;
RA Rebay I., Chen F., Hsiao F., Kolodziej P.A., Kuang B.H., Laverly T.,
RA Sub C., Voas M., Williams A., Rubin G.M.;
RT "A Genetic Screen for Novel Components of the Ras/Mitogen-Activated
RT Protein Kinase Signaling Pathway That Interact With the yan Gene of
RT Drosophila identifies split ends, a New RNA Recognition Motif-
RT Containing Protein."
RL Genetics 154:695-712(2000).
DR EMBL; AF184612; AAF26299.1; -.
DR HSSP; P09651; 1HA1.
DR FLYBase; FBgn0016977; spen.
DR InterPro; IPR000504; RRM.
DR SMART; SM00360; RRM; 3.
DR PROSITE; PS50102; RRM; 3.
SQ SEQUENCE 5476 AA; 590531 MW; 93FAA8C7860770C2 CRC64;

Query Match 4.48; Score 334; DB 5; Length 5476;
Best Local Similarity 20.36; Pred. No. 6.9e-11;
Matches 331; Conservative 197; Mismatches 654; Indels 450; Gaps 74;

Oy 8 TSDSRATRRKRCPPDGLGSPKNTKRRNREQNKTEELAEILFANFNIDIONFNKP 67
Db 2762 TSAPSTQTSKRG-----EDMEPTFGIISDEESQF-P 2794

Oy 68 DKCALKETVROIKOEKRAAANIDEVOKSDVSTGGVYIDKALGPMLEALDGF 127
Db 2795 EGA---ETNK-----DILPSSYST-----GPIYALQ--- 2820

Oy 128 FVNLEGNVFESENVTOYLRYNOEELMNKSVSYLHV--GDHTEFVKNLPLKSLVNGS 185
Db 2821 -----TYKQEPSTPNKNEBAHIQLTVNEPEQOOOLERSRLSGS 2860

Oy 186 WSGEPRRNSHTFNCBMLVKPLRPOSEEGHNDQEAHQETMQCAVYQPSKI--KEGE 243
Db 2861 SSSSHADERIRRRERK--EKRRREKSOREQOQ--IHQSSKET--KVDDNSVDMDEGR 2916

Oy 244 DLQSLICVARRVPKERPVLPSSSEFTTRDLOGKITSLDTSTRAAMKPGMEDLVRC 303
Db 2917 ALAQL-----MSDDTFRPIS-----EATPRTAATYRSMDT----- 2949

Oy 304 IQKPHAGHEGSVYAKRHHHEVLFQGLAFSQTFRFSLSDGLVAQAQTKSKILRSQTINE 363
Db 2950 VFRSEDNNDNSV-----DMTKQGVKSEQEQHKKSKD-----KKKKRKRKEERQ 2994

Oy 364 POLVYSLMLAREQNVCMNPDLTGOTMGK-----PLNPISSPHQAOLC 409
Db 2995 EKL---LQOORRESLPNVAISSAPPTPGKLTVNQAASKNADLDLDAKHISSP---PVC 3048

Oy 410 SGNPQDMLTSSNINPFRNGREQMGMPRGFGSGGNHVSQGAOTTPGGSNTVALKMS 469
:
```

```

Db 3049 KPSPSLPCLIGDDDDDLATHTPAKRTPPSSR--GNDGL-----T 3085
Oy 470 PSSQSPGMNQGPPTSMLSPRHRMSPGVAGSPRIPSPQSPAGSLHSPVCGSSTGNHSY 529
Db 3086 PSREKP-----RLISP-----IKPTPIANSSTLTSTQSAETTPV-----SGGTVIS- 3125

Oy 530 TNSLSNALQALSEGHVSLGSSLASPDLMGNLQNSPVNMMNPPLSKMGSLSDKDCF--G 587
Db 3126 -SSALATPTSTTAGVS-----AAP-----GLDMSPT-----SASAQCKKRSFIPG 3167

Oy 588 LYGEPESETTGOAESCHPGEOKET-----NDNLPRAVSSERADGOSRLHDSKGQTK 640
Db 3168 FDGQIDDDRISSAVQSI--SAEFSNSTLNDIADDEKRIIVASPRATKPLDKRIEESKRYT 3226

Oy 641 LLQLLTTSQDMEPSPLASLSDTKDKDTSGLPGSGSTHGTSLKKEKHILHRLDSSSP 700
Db 3227 ISQ-----EETESAVSALGSEFTGSTTDTYSLDGDEKSVNELTPPL--VINEPDEE 3279

Oy 701 VLAKLTAEATGKDLIS-----QESSSTAPGSEVTIKQEPV-----SPKKKENALLRY 747
Db 3280 AALAKAIETAGERASILEEPREMERAREPDPDEALIESPPVEVLDPELNKAV--Q 3337

Oy 748 LIDKDDTKDIGLPEITPKLER--LDSTDPASNTKLIAMKTEKEMSFEPGDDPGSELD 804
Db 3338 SLKHEDMDIKAD--TPQSERDQIDTDEENPDEADSSGSLKIDETVQSSSSPEKSIS 3395

Oy 805 NIEELLDLQNSQLPOLPRPRPGAPAGSVKQALINLMQLTANSPVT--PVGAQKXAL 863
Db 3396 NNSPTPRETANIDIPNV--ESQPKLSNESTPPQSVITLPLF--DTPRTVAGLPPSPV 3450

Oy 864 RISQSTFNN-----PRPGQLGR--LLPQONLPLD----- 890
Db 3451 KIEPRTISKLOQPLVQPVQTVLPAHSGSGISANSVINLDSNVISSGCSNTSASATAS 3510

Oy 891 ----ITLQSPFGAGPPPIRNSSPYSVTPQGMGNGCMIGNQMGILNCSSTGMICNSASR 946
Db 3511 ASASISFSGSPITAS-----QNAMPQASTPKQGPPIRQQAIRFQSLIMOPTPISIEQTPH 3564

Oy 947 PTMPSGEMAPQSAVRVCAATSAAMNRYVQGM-----IRNPAASIPRPSQPGQORTL 1002
Db 3565 FAVPQMWLSPOS-----HNPQQGTVMGVRAS--PSPSLSPKRGVAAQ 3607

Oy 1003 QSOVANNIGPSELEMMNGRFOYS-----QQQAPNQ-----TAPMESELPIPOAS--PAS 1050
Db 3608 SRLVQGLSP-----VGRPMYSQPSPOQOYQOTQOQHALLITSPOSSNISPLASPTTRYLS 3661

Oy 1051 QNRQPGS-----SPDILCP--HPAAESPDEGALLDLULYALRNPDLGLEIDRALGPE 1104
Db 3662 SSNSPTTSKVNSYQPRNOQVPOQPSKVAEVOTTPOLMTPIPLQKMTPIQVPHN-----PT 3717

Oy 1105 LVSQSOAVDPDEFSSOD-----SNIML-----EOKAVYFPOQYASQAQMAQ 1146
Db 3718 IISKVVTVQPOQATQSOVASSPPLGLSPRHKNVNLNANQNOQOPVIAKMTAHOQHOHQ 3777

Oy 1147 SY-----SPMDPNEHTGWRSPYATLTMORPRGLRPTGLVQNOP--NQLRL--- 1191
Db 3778 QFMHQQMTQROQHMQOQOLHGQSQDITSA-----FOHQMHQOHOQOQOHHNQHNLQO 3832

Oy 1192 -----QLOHRLQAOQNRQPLMQNISVSNVNLTLRGVTPQARINAMQLAQORETL 1243
Db 3833 LHAQOHPTQKHQAOQOQFNOQLOQHQSOU-----QHVOYQOQAOQOHNLSQ 3878

Oy 1244 NQHLQROQHQQOQYQQRTL-----MKRG--QGLMTPRSMVAPRG-----MPATMSNPR 1290
Db 3879 QOHQSOQOQLOQHOQAOQOQLOQIQKLOQHNGPQOQKSPQVGHGIGSGSTIFASQOHNHQ 3938

Oy 1291 IP-----QANAQGF-----PEPPNYGIS-----QQPDPGFTGATTPQSPILMSPRMHTOS- 1335
Db 3939 LPARGVPOQHHPQOLSHSPCKRNTLVNSQVQGRPALITRGHSISQPNQOQOQLHQOSS 3998

Oy 1336 ---PMQOQSO---ANPAYQAPSDINGMAQGNKGSMSFQSQSPPHFGQOANTSMYSNNKN 1389
Db 3999 SGHPRHQKOLSSPGANILPQDTPLNVIQNTPKLIVQOHIVAQNOQVPPQGTGNAIHYPQNG 4058
```

QY 1390 INVSMATNGMGSSNMOTGOISMTSV-----TSVSTISGLSSMGPEQVNDPALNGNIFPN 1445  
 Db 4059 KD---STPPGHVPTTPMAKSQTSSESVIRTPPTGLAVISANTVGS-LTTEENLIK 4114  
 QY 1446 QLPGM-MIKOE 1456  
 Db 4115 SOPKODELEED 4126

RESULT 26  
 09VPL2 PRELIMINARY; PRT: 5533 AA.

AC 09VPL2; 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE SPEN PROTEIN  
 GN SPEN OR CG18497.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_Taxid:7227;  
 RN 111  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BERKELEY;  
 RX MEDLINE-20196006; PubMed-10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abrial J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,  
 RA Burdits K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K.J., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Foster C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,  
 RA Glodde A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harits N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Mlshina N.V., Mobarry C., Morris J., Moshirei A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclib J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden Klamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svitskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Maasarm D.A., Weinstock G.M., Weisenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster.";  
 RL Science 287:2185-2195(2000).  
 DR EMBL: AE003590; AAF51534.2; .  
 DR HSSP: P09651; IHA1.  
 DR FLYBASE; FBgn0016977; spen.  
 DR InterPro; IPR000504; RRM.  
 DR Pfam; PF00076; rrm; 3.  
 DR SMART; SM00360; RRM; 3.  
 DR PROSITE; PS50102; RRM; 3.

SQ SEQUENCE 5533 AA; 597145 MW; 713E21117EAF2CCD CRC64;  
 Query Match 4.4%; Score 334; DB 5; Length 5533;  
 Best Local Similarity 20.3%; Pred. No. 7e-11;  
 Matches 331; Conservative 197; Mismatches 654; Indels 450; Gaps 74;

QY 8 TSDPSRAETRRKRECPDOLGSPSKRNTKRNROENKYEIELAELIFANFDIDNFKP 67  
 Db 2819 TSAPSTKQTSKRG-----EDKMEFTFGIISDEESQF-P 2851  
 QY 68 DKCAILKETVKQIRKREKKAANAANDEYOKSDVSTGSGVIDKALGPMLEALDGF 127  
 Db 2852 EQA-----ETNK-----DIIPSSVST-----GPVSAALQ--- 2877  
 QY 128 FVYNLEGNVFVSENVYQLRYNOEELMNKSVSLHV---GDHTEFVKNLIPKSYNGS 185  
 Db 2878 -----TYKQEPSTPKSKMEAHILQTLVHEPEQOOLERSRLSGS 2917  
 QY 186 WSGEPPRRNSHTNCRLVPLPDSEEGHNOEHOKEYTMOCFAVSOPKSI-KEEGE 243  
 Db 2918 SSSSHADREHRRREKR--EKRRREKSQREQNQ--HQKSKVET-KYDDONSVDMDQAGR 2973  
 QY 244 DLOSLICVARRVPMKERVLPSSSEFTTRQDLQKTTSLDTSTMAAKPGMEDLYRC 303  
 Db 2974 ALERQL-----MSDFTKPISE-----EATPSTATYRSDMTD----- 3006  
 QY 304 IQRFHQHEGESVYAKRHNEHVLROGLAFSQIYRFSLSGTLVAQTKSLIRSQTTNE 363  
 Db 3007 VRFPSDNEDNNSV-----DMTKQCVKSEQDQHKSKD-----KKKKRRSKREKO 3051  
 QY 364 POLVLSHMLHRBQNVCMNPDLTGQTMK-----PLNPISNSPRAHQALC 409  
 Db 3052 EKL-----LQQRRLSLPVAISTASAPPLPKGLTVNVOAKSHADLDQAKHISP---PVC 3105  
 QY 410 SCNPQODMTLSSNINPFPNGKMGKPRFGSGGMNIVSGMOTTPGSGVYALKMS 469  
 Db 3106 KPSPSLPCLIGDDDDALHTPKAKPTTPSSR--GNGL-----T 3142  
 QY 470 PSQSPGMNPGQPTSMILPBRHMSPGVAGSPRIIPSPFSPAGLSHPVGSSGTNSHSY 529  
 Db 3143 PSREKP-----RLSP-----IPKPTIANSSTLSTQSAETPV---SSGVIYS- 3182  
 QY 530 TNSLNAQLALSGHGVSLGSSLASPDLMKGNLQNSPVNNPPLSKMGSLGSKDCF--G 567  
 Db 3183 -SSALATTPPTSPYAAQVS-----AAP-----GLDNSPT-----SASACKKKESEFIPG 3224  
 QY 588 LVGEPSGCTTGOALESSCHPGEOKET-----NDPNLPRAVSSERRADGGRILDSKQTK 640  
 Db 3225 FDSQLDRLSESVAQSI--SAEFNSTSLDNIADPEKIPVAPPRATKPLDKLEESKSRVT 3283  
 QY 641 LLDLLTTKSDQMEPPLASSLSDTNKDSGSLPGSGSTHGTSLKEKHLIHRLELSSSP 700  
 Db 3284 ISQ-----EEFESAVNALGSPFTSSTTDVSLDGMEMSVNLEPTTL---VIAEDDEE 3336  
 QY 701 VDLAKTLAEATQKGLS-----QSSSTAGSEVYTIKQEPV---SPKKEMALLRY 747  
 Db 3337 AALAAAIETAGPASTLEPEMEPREAREPDDPEALEISEPVEVLEDEELKAV--Q 3394  
 QY 748 LLDKADTKDGLPELTPKLER---LDSKTDPAENTKIAMKTEKEMSPERGQPSGLD 804  
 Db 3395 SLKHEDMDIKAD--PPQSERDQIDITVEENDEADSSGSLKIDETVQSSSSPESKIS 3452  
 QY 805 NLEELIIDDQNSLOLPUPPTPCGAPAGASYDKOAIINDLMQLAENSPPV--PVGAKTAL 863  
 Db 3453 NNSPTRETRANDIPNV--ESQKLSNESIPQSVITTKLPFL---DTPKTPAGLPPSPV 3507  
 QY 864 RISQSTFNN-----PRPOLGR-LLPNOLPLD----- 890  
 Db 3508 KIEPPTISKQGLPVLQPVQVTLVLPAPHSSTGSIANSVINLDSNVSSTSAASATAS 3567  
 QY 891 ----ITLQSFSGAPPPPIRNSSPYVITPQPGMNGMGIMGNLGNSTGIMGNISASR 946

```

Db 3568 ASASISFGSPAS-----ONAMPDASTPKOGRIPPOQAIRTOSLIMOPTISIEQTPH 3621
Qy 947 PTMSEGAPOSASAVRYTCATTSAMNRPVOGGM-----IRNPASISMRPSSOPEGOROTL 1002
Db 3622 FAVPOMVILSPS-----HHPOOPGYMVGIRAPDS---PSPHSPERGVAO 3664
Qy 1003 OSOVMNIGSPSELMNMGSPOTS-----OOQAPPNQ-----TAPMESTILPIDOAS--PAS 1050
Db 3665 SRIVGOLSP-----VGRPMVSQSPPOQVOOTQOQHALITSPOSNTSPPLASTPTRYLS 3718
Qy 1051 QNRPPGS-----SPDILCP-HPAESPDEGALLDQLYALRNFDGLEETDRALGPE 1104
Db 3719 SSNSPTTSKVSXOYPRNOQVPOQSPKSVAEVOTPTQMTPLDQMPTEIYVPHN-----PT 3774
Qy 1105 LVSOQANDPROFSQD-----SNIML-----EQKAPVFQOYASQAMQAG 1146
Db 3775 IISKVYVYQPOQATQSOYASSPRLGSLPRPHKNVHLNANOQOOPQVYAKMTAHOHQHMQ 3834
Qy 1147 SY-----SPMDPFTMGORPSYATLIMQPRPGLRPTGLVQNOP-----NOLRL--- 1191
Db 3835 QFMHQMTORQOHMQOOLHQSQOITSA-----FOHQMOHQOAOQOQHNNQOHLMOQ 3889
Qy 1192 -----QLOHRLQAOQNRQPLMNOISVSNVNLTLRPGVPTQAPINAMLAQOREIL 1243
Db 3890 LHAQOHPTQOHOAOQOQENQOIQOHSQO-----QHOVOQOAOQOQOHLISQ 3935
Qy 1244 NOHLRQRMHOOOYOQRTL-----MMRG-QGLMTPSMVAPBSG-----MPATMSMPR 1290
Db 3936 QOHOQOQOOLNOHOAOQOQOIOLOKLOQMHGPOOQKSPQVGHGIGSTSIFFASQOHSQ 3995
Qy 1291 IP-----QANAQOF-----PPPNYGIS-----QOPDPFGTATTPQSPILMSPRMAHTOS- 1335
Db 3996 LPAQVPOQOHPROOLSHSPCKPMTIVSVNGVOPRALITRVGHSIQNOOQOLPHQOSS 4055
Qy 1336 ---PMMQOSO---ANPAYQADSDINGMAQMGNSMFSQOSPHEGQOANTSMYSNNMN 1389
Db 4056 SGHHQKOLSSPGANLPLQOTPLVNIQMTPKIIVQOHIVAQNOVPPPOGNAIHYPQNOG 4115
Qy 1390 INVSMATVTGMSMNQMTGQISMTSV-----TSVTSGLSSMGPRQVNDPALRGNLFPN 1445
Db 4116 KD---STPPGHVEPPAPASAKTSBSVSIVIRPTPTGLAVISANTVGS-LITEENLIKI 4171
Qy 1446 QLPQMD-MIKOE 1456
Db 4172 SQPKODELIBD 4183

```

## RESULT 27

```

09NHN1
ID 09NHN1 PRELIMINARY; PRT; 5554 AA.
AC 09NHN1;
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE SPLIT ENDS LONG ISOFORM.
GN SPEN OR CG18497.
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxId=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Huang B., Wu S., Shin Y.-A., Luo L., Kolodziej P.;
RT "Split ends encodes large nuclear proteins that regulate neuronal cell
rate and axon extension in the Drosophila embryo."
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF221715; AAF34661.1;
DR HSP: P09651; 1HAL;
DR FlyBase; FBgn0016977; spen.
DR InterPro; IPR000504; RRM.
DR SMART; SM00360; RRM; 3.
DR PROSITE; PS50102; RRM; 3.

```

```

SQ SEQUENCE 5554 AA; 59918 MW; 4037E27833D0C622 CRC64;
Query Match 4.48; Score 334; DB 5; Length 5554;
Best Local Similarity 20.38; Pred. No. 7.1e-11;
Matches 331; Conservative 197; Mismatches 654; Indels 450; Gaps 74;
Qy 8 TSDPSRAETRRKKECPDGLGSPKNTKERRRNEQNKYTEELAEILFANFNIDIDFNKRP 67
Db 2813 TSAPSTQTSKRG-----EDMEFTFIIISDEESQF-P 2845
Qy 68 DKCALIKETVQRIQIKOEKAAANIDEVOKSDVSYSGQGVYDKDALPMLLEALDFF 127
Db 2846 EQA-----ETNK-----DILPSVST-----GPIVSALQ--- 2871
Qy 128 FVNINLEGNVFESENTOYLRYNOELMNKSVYSLHV--GDHFEVKNLLPKSLVNGS 185
Db 2872 -----TYKQEPSTPNKNEEAHQLIVHEPQOQOOLERSRLSGS 2911.
Qy 186 MSGEPRRNSHTFNCRMLVYKPLPDEEGHDNOEAHQYETMQCPAVSQPSI--KEEGE 243
Db 2912 SSSSHADERRRRRERK--EKRRKRSQREQNO--IHQKSKVET--KVDNDSVDMDEAGR 2967
Qy 244 DLOSCLICVARRVPKKEPVLPSSEFTTRDLOGKITSLDTSTRAAMKFGMEDLVARC 303
Db 2968 ALEAQL-----MSDETFRPIS-----EATPSTAAVTRSDMTD----- 3000
Qy 304 IQKFAHQEGESVYAKRHHHEVLRQGLAFSQIYRSLSDGTLYAAQTKSLINSOTTNE 363
Db 3001 VEFESDNDNNVS-----DMTKQGVKSEQEQEHKSKD-----KKKKKKRKEEKQ 3045
Qy 364 POLVISTLMHLEQNVCMNPDLTGQTMK-----PLNFSSNSPAHQALC 409
Db 3046 EKL---LQOQRESLEPNASTSSAPPTPGKLTVNVQAASKHADQLDKKHITSP--PVC 3099
Qy 410 SGNPGQDMLTSSNINFPINGPKQMGMPRGFGSGGNHVSQMATTPQGSNTYALKMS 469
Db 3100 KPSPLPLCLIDDDDLAHTKAKTTPSSR--GNDGL-----T 3136
Qy 470 PSQSEFGNPGQPTSMLSLRHMSFGVAGSPRIIPSOFPSPAGLSHPYGVCSSTGNSISY 529
Db 3137 PSREKP-----RLISP-----IPKPTIANSSTLSQSAETPV-----SSGTVIS- 3176
Qy 530 TNSSLNALQALSEGHVGLGSSIASPDLKMGNLQNSPVNMPPLSKNGSLDSKCF--G 587
Db 3177 -SSALATTPSTTAGVS-----AAP-----GLDSTP-----SASAQCKKKSFTPG 3218
Qy 588 LYGEPESECTGOAESCHPGEOKET-----NDNLPNAVSEERADQSLRHDKSGQTK 640
Db 3219 FDGQLDDRLISSAVQSI--SAEFNSTSLDNTADEKTIIVASPPRATKRLDKLEBSKRVY 3277
Qy 641 LIQLITTSQDMESPPLASSLSDTNKSDTGSLSPGSGSTHGTSLEKHKILHRLLODSSP 700
Db 3278 ISO---RETESAVSALIGSEFGTSTTDYSLGDMEKSVNELETPTL---VIAPEDEE 3330
Qy 701 VDLAKLTAEATGKDS-----QESSSTARPGEVITIKQEPV---SPKKKEALLRY 747
Db 3331 AALAKALETAGEPASILIEBPMEPERAEPPDEADEIESEPVEVLDPRELKNAV--Q 3388
Qy 748 LLDKDDTDIGLEITPKLER---LDSKTDPASNTKLAMTEKEKMSFEFGDQGSGLD 804
Db 3389 SLKHEDMDIKAD--TPQSENDLQIDTTEENPDPADSSGSLKIDETVQSSSEKKSIS 3446
Qy 805 NIEETILDLONSQLPFLPDPTRPGAPAGSVDKAIIINDIMOLJFENSFVT--PVGAQKTAL 863
Db 3447 NNSPTPRETANIDIPNV--ESQPKLSNESTPOPSTVITKLPLV--DTPKTVPAGLPSPV 3501
Qy 864 RLSOSTENN-----PRGQLGR--LLPQNQLPLD----- 890
Db 3502 KIEPTISKLOQPLVQVQTVLPAPHSVGSISANSVINDLSNYSISSCNTSAASATAS 3561
Qy 891 ---TTLSPTGAGFPPIRRNSPYSVLPQGMNGQMINQGNLGNSSGMIGNSSAR 946

```



```

Db 3562 ASASIFSGSPTAS-----QNMPOASTKRCGGITPQOAIKTOSLIMQPTTISREOTPH 3615
OY 947 PTMPSGEMAPSSAVRYTCATTSAMNRPVQSGM-----IRNPAASIPMRSPGRCORLT 1002
Db 3616 FAYPQAVLSPQS-----HHPPQRPETYWVGLRAPS---PHSLHSPRGVAG 3658
OY 1003 QASOVNMGISELEMMNGGPOYS-----QQCAPRPO-----TAPWPSILPTIDQAS--FAS 1050
Db 3659 SRLVQGLSP-----VGRPVWSQPSPOQOQOQOQHALLTSPQSSNISPLASPTTIVLS 3712
OY 1051 ONKPRGSS-----SPDLLCP--HPAESPDEGALLDOLYALRNPDGIEIDRALGIPE 1104
Db 3713 SSMSPPTSKVNSYQPRNQVPOOPSPKSAVEVOTPOLMTPILOKTPLOVPH-----PT 3768
OY 1105 LVSQASAVDEPESD-----SNIML-----EOKAPVFPQOYASQAQMAOG 1146
Db 3769 IISKVTVPOQCATQSVASSPPLGSLPRPKVHLNANHQOQOQOYLAKTANHQOQHM 3828
OY 1147 SY-----SPQDPNHTMGQRPSTYATLRMQPRRGLRPTGLVONP---NQLHL--- 1191
Db 3829 QFWHQQMIQROQHMOQOOLHGOQOQITSA-----POHMQOQOQOQOQHNOHNLNOQ 3883
OY 1192 -----OLOHRLQAOONRQPLMNOISVSNVNLTLRPGVTPQAPRLNQMLAQOREIL 1243
Db 3884 LHAQOHPQOKOHAQOQOQFNQIQOQOQO-----OHQOQOQOQOQOQOHLNO 3929
OY 1244 NQHLRQROMHQOQOQOQOQOQOQOQOQO-----MMRG--OGLNMTPSMAVPSG-----MPATMSNPR 1290
Db 3930 QOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQO 3989
OY 1291 IP-----QANAOQF-----PFPNYGIS-----QOPDGFGLGATTPQOQOQOQOQOQO 1335
Db 3990 LPARGVPOQOHPQOLSHSPCKPNTLVSNQVQVPPALILRVSHSQPNQOQOQOQOQOQO 4049
OY 1336 ---PMQOQSO---ANPAYQAPSDINGMAQMGMSQSPHFGQOQANTSMYSNNM 1389
Db 4050 SGHPHOKOLSSPCANLPLQPLVNIQNTKTIQOQHVQOQVPPROTQOQNAIHQONOG 4109
OY 1390 INVSMAINTGMSNMQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQO 1445
Db 4110 KD---STPQGHVPTPAMSAQKTSSESVYRTPTTGLAVISANTVGS--LTREENLIK 4165
OY 1446 QLPQMD-MIKOE 1456
Db 4166 SQKODELIED 4177

```

## RESULT 28

```

OYVPLI PRELIMINARY: PRT; 5560 AA.
AC Q9VPLI:
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE SPEN PROTEIN.
GN SPEN OR CG18497.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxId=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Branton R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Abdayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,

```

```

RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Giodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Haris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibeigwan C.,
RA Jalili M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Mlshina N.V., Mobarry C., Morris J., Moshirei A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclebe J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Palzer K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Spler E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svitkas R., Tector C., Turner R., Ventler E., Wang A.H., Wang X.,
RA Wang Z.-Y., Nassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Zheng H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL: AE003590; AAF51535.2; -.
DR HSSP: P09651; 1HAL.
DR FLYbase; FBgn0016977; spen.
DR InterPro; IPR000504; RRM.
DR Pfam; PF00076; rrm; 3.
DR SMART; SM00360; RRM; 3.
DR PROSITE; PS50102; RRM; 3.
SQ SEQUENCE 5560 AA; 599996 MW; 58F19621AF40D2A8 CRC64;

```

```

Query Match 4.4%; Score 334; DB 5; Length 5560;
Best Local Similarity 20.3%; Pred. No. 7.1e-11;
Matches 331; Conservative 197; Mismatches 654; Indels 450; Gaps 74;

OY 8 TSDPSRAETRRKKECPDQGLPSPKRTKRNROEKNYIEELIFANFNIDNENFRP 67
Db 2819 TSAFSTAQTSKRG-----EDKKEFLFGIISDESEGF-P 2851
OY 68 DKCALIKETVQKIQKEQKAAANIDEVQKSDVSTGQVLDKDALGPMMLLEALDGEF 127
Db 2852 EQA---ETNK-----DIPSSVST-----GPIVSALQ--- 2877
OY 128 FVYNLEGNVVFSENTQYLRYNOEELMKSVYSILHV---GDHTEFVKNLDPKSYNGGS 185
Db 2878 -----TYKQEPSTPNKNEAHIDLVNEPQOQOQLERSIRLSGS 2917
OY 186 WSGEPRRNSHTFNCBMLVPLRPSDEEGHNDQEAHQKYTECFVAPSPKSI--KEEG 243
Db 2918 SSSSHADRENHREKR--EKKRREKQOREQON--INOKSKVBT--KVDDNDSVMDQEAR 2973
OY 244 DLOSCILCVARVYMKERVLPSSESFTTRDLOGKITSLDTSTMAAKPMGEDLVRR 303
Db 2974 ALFAOL-----MSDFQTKPRISE-----EATPSTATVYDSMDTD----- 3006
OY 304 ICKFNQHEGESVYAKRHNNHVLROGLAFSQIYRSLSLDGTLVAQOTSKLIRSQTNE 363
Db 3007 VEFPSDNEDNSV-----DMTKQGVSEQOQOQOQOQOQOQOQOQOQOQOQOQOQOQO 3051
OY 364 POLVILHMLHRRQNVCMVNPDLTGQTMGK-----PLNPSSNSPAPQALC 409
Db 3052 EKL---LQOQRSLRSLNASTSSAPRTPKGLVYNVQAASKHADLDQDAHHSPP---PVC 3105
OY 410 SGNPQDMTLLSINPFLNGPKQKMPGMRGREGSGGMNHVSGMAVTPQGSNYALKMS 469

```

```

Db      3106 KSPSLPCLIGDDDDALHTPAKPTTSSR--GNDGL-----T 3142
Qy      470 PQOSSPGMPPGPTKSLSRHRHMSPEVACSPRIPPSQEPAGSLSPVVCSTGNSHY 529
Db      3143 PSREKP-----RLISP-----IPKPTIANSSTLSTOSAEPTV-----SSGTVIS- 3182
Qy      530 TNSSNALQALSEGHVSLGSSSLASPDLMGNLONSPVMMNPPLSKMKSIDSKOCF--G 587
Db      3183 -SSALATPTTSTTAAGVS-----AAP-----GUDNSPT-----SASACKKKESTFPG 3224
Qy      588 LYGEPESEGTGQAESCHPEQKET-----NDPNLPVAVSERADGOSRLHDSKQTK 640
Db      3225 FGGQDDRRISASAVQSI--SAFNSTSLDNINADPKIPAPSPRATKPLDKLEESRAT 3283
Qy      641 LLQLLTTSKDOMPEPPLASSLSDTKNDSTGSLPGSGSTHGTSLKEKHKLHRLLODSSP 700
Db      3284 ISQ-----EETESAVSALLDESFGTSTTDYSLDGMDEMSVLEETPTL---VIAEPDEE 3336
Qy      701 VDLAKLTAEATGKDL-----QESSSTAPEGSEVTTKOBPV-----SPKKKENALLRY 747
Db      3337 AALAKALETAGEPASILEEPERERAEPPDPAETSEPEVVEVLDPEELNKAIV--Q 3394
Qy      748 LLDKDDTDIGLPEITPKLER--LDSKTDPAANTKLIMKTEKEEMSEPEPDQPSELD 804
Db      3395 SLKHEDMDIKAD--TPOSERLQIDTDEENPDEADSSGSLKIDETVQSSSPERKSTIS 3452
Qy      805 NLEELILDIONSQLPOLFPDTPPGACAGVCKQAIIINDIMOLTAENSPVT--PVGAOKTAL 863
Db      3453 NNSPTPRETANIDIPNV--ESOPKLSNESTPOPSVTYTKLPFL--DTPKTVAGLPPSPV 3507
Qy      864 RISQSTFNN-----PRPGQLGR--LLPNONLPLD----- 890
Db      3508 KLEPPTISKLOOPLVOPVOTVLPABHSTGSGISANSVINLDSLNISSCSNTSASATAS 3567
Qy      891 ----TLOSPTGAGFPPIRNSSPSVIIPQPMGNGMIGMGNLGNSTGMIGNSASR 946
Db      3568 ASASISFCSPTAS-----QNMPPQASPTPKQPIPPQQAIRQSLIMOPPTISIPQPTH 3621
Qy      947 PPMPEGMAPOSSAVRYTCAATTSAMNRPVCGM---IRNPASTPMRPSPOGQOROTL 1002
Db      3622 FAVPMVLSPOS-----HHPOQPGTYWVGIRAS---PHSPHLSPGKGVAVQ 3664
Qy      1003 OSQVANNIGSELEMMNGGPOVS---OOQAPRNO-----PAWPESILPIDQAS--FAS 1050
Db      3665 SRLVQOLSP-----VGRPMWSQPSPOOVQOOTQOOHALITTSPOSSNISPLSPTRVLS 3718
Qy      1051 QNRQPFGS-----SPDDLICP--HPAESPSDEGALLDQLYLALRNFDGLEIDRALGIRE 1104
Db      3719 SENSPTTSKVNSTYORPNQOVPQPSKSAVEVOTTPQMLTIPLOKMTPIQVPHH-----PT 3774
Qy      1105 LVSQSQAVDPEQFSSOD-----SNIML-----EOKAPVFPQOQVLSQAOMAG 1146
Db      3775 IISKVVTYVOPQATOSVAVSAPPLGSLRPKHVNHNAHONOOQOPOTIATKTAHONQHMQ 3834
Qy      1147 SY-----SPMDPNHTMGORPSVATLRLMORPPGLRPTGLVONOP--NOLRL--- 1191
Db      3835 QPMHQMOTOROHMOOQOLHGSOOTSTA-----FOHMHQOHQAOOQOHNNHNOOHLNQ 3889
Qy      1192 -----QLOHRLQAOONRPLMNOISVNVNLTLPBPGVTOARINAMLOAORELT 1243
Db      3890 LHAQOHPTOKHOAOQOQFNQOIQOHQSOQ-----QHQQVQOQNAOQOCHLSQ 3935
Qy      1244 NOHLROROHMOOQOQOQRTL-----MMRG--QGLMTPRSMVAPSG-----MPTMSNPR 1290
Db      3936 QOHQSOQOOLNDOHQAOQOQLOQIQKLOHMHGPOQOQKSPQGVGHGSGTISIFASQOHNQ 3995
Qy      1291 IP-----QANAQOF-----PFPNTYGIS-----QOPDPGCTGATTPQSPMSPRMAHTOS- 1335
Db      3996 LPARGVPOQHPOQLSHSSPCPKPNTLVSVNQGVOVPAITTRGSHSOPNQOQOLPHQOSS 4055
Qy      1336 ---PMHQSOQ---ANPAYQAPSDINGMAQNGNGSMFSQSQSPPIHGQOQANTSMYSNNNN 1389

```

```

Db      4056 SGHPHOKQLSPGKANLPLOQPLNVIONTPKIIYQOHIVAQONVPPQOTGNAIHYQONG 4115
Qy      1390 INVSAMTNTGMSNMOMTGOISMTSY-----TSVTSIGSLSSMGPEVYNDPALRGULFPN 1445
Db      4116 KD---STPGHPEPPPMASQKTSSESVIRKPTPTTGLAVISANTVGS-LTLENTLIRI 4171
Qy      1446 QLPQMD-MIKOE 1456
Db      4172 SQPKDELTIED 4183

RESULT 29
09V86 PRELIMINARY; PRT: 2280 AA.
ID 09V86
AC 09V86
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE CG5580 PROTEIN.
GN SBB OR CG5580.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxId=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY.
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA Anatolitsas S.E., Binkley S.E., Blakes R.G., Champs C.R., Pfeiffer B.D.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borrova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Cantler A., Chandra I.,
RA Cherry J.M., Crawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos J.B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K.J., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferlita S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Giodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Mei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Maltel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclab J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Relbert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E.C., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svitskas R., Tector C., Turner R., Venter G.E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao O.A.,
RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
EMBL: AF003800; AAF5721.1;
DR Flybase: FBgn0010575; sbb.
DR InterPro: IPR003880; Znfphopant_attach.
DR InterPro: IPR000822; Znf-C2H2.
DR Pfam: PF00096; zf-C2H2; 1.
DR SMART: SM00355; Znf_C2H2; 1.

```

DR PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN\_1.  
 DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 1.  
 DR PROSITE; PS50157; ZINC\_FINGER\_C2H2\_2; 1.  
 KW DNA-binding; Metal-binding; zinc-finger.  
 SQ SEQUENCE 2280 AA; 228107 MW; 1C2CDA27DC77B00B CRC64;

Query Match 4.4%; Score 332; DB 5; Length 2280;  
 Best local similarity 20.6%; Pred. No. 2.6e-11;  
 Matches 266; Conservative 138; Mismatches 452; Indels 438; Gaps 59;

QY 426 PINGKQMGMPKRGSGSGMHNVS-----GMQATTPOG-----SNY 463  
 DB 1017 PLMGDASD-GASVGGIGAGGVNNAAGIPISASGGGLATQPPSLNPVTGLNVOISTK 1075  
 QY 464 ALKNNSPSSQSPGM--NPGOPTSM-----LSPRHMSPGVAG-----SP 500  
 DB 1076 KCKTASPCAISPVLLCEPDDCCSKYKHNGLRYHOSHAGAGGASSMDEDSQAPEDP 1135  
 QY 501 RIPSPQSPAGSLHSPYGVCSSTGNSHYTNSLNLALQALSEGHVSLGSLASPDLKMG 560  
 DB 1136 ATPSP-----GVASGTGSGASVASAVPA-TAPSAQ-----GVAVSPNTPLA 1179  
 QY 561 NLQNSPVNMM-----PPPLSMGSLDSKDCGLY 589  
 DB 1180 NSSNPVTNGVNAAPSAPATGVTITAPNTTPSVVETQAPLTGPPEVT----- 1225  
 QY 590 GEPSEGTGGAESCHGDEKETND-----PMLPRAVSSERADGGSLHDSKGTKL--- 642  
 DB 1226 --PAAPPTICAVAT--PGAEOVSIVPLGLGLPLTAGPNSGTQQQPPPTQQQPOLLVPG 1281  
 QY 643 -QLLTTRKSDMEPSPLA-----SLSLDTNKDSTGSLPGSGSTHGTSLKEKRLH 691  
 DB 1282 GSAASLIQ00000QPVAGSGITAGISQALSLQHQQQLMGLP-----AMLSDQQ--Q 1332  
 QY 692 RLDDSS-----SPVD--LAKLTAETGKDLSESSSTAGSEYTKQEP----- 734  
 DB 1333 ALLQ0GALKAGVLFNRPDPDNPLOQ0GQASVNPQ0QSPPRP--SHVQDQTPSAYAQ 1391  
 QY 735 -----VSPKKEN-----ALLRYLDDKDDTKDIGLEI---TP--KL 766  
 DB 1392 AGLKTSFGFSGVGAASSKOKKRRKSPGSPDFRGVSRREVOGSPAYSDISDSTPAE0 1451  
 QY 767 ERLDSKTDPAASNTK--LIAMKTEKEMSEPPDQSGSELDNLEILLDLONS----- 816  
 DB 1452 EMLDKSVGAQVATKATIELMGKKRPEVGVPPAPPMNVYPMQGFYPAQ0QASAPRQ0Q 1511  
 QY 817 -QLPQLPDRPRGAPASSVDK0AIIINDLMQITANSPT-----PVGAKTALRIS 866  
 DB 1512 Q0Q0FYVQTERGKPRGLPALTQA0Q0Q0LQRCAPRPTSQRPSHLLGPRQ0QSVAAHLA 1571  
 QY 867 QSTFNN-----PRPG-----QLGRLLPQNQLPLDI-TLQSPGAGPPPIR 906  
 DB 1572 DYSCKNNDPRLDMLTKRQRPQRPQ00Q0SGLSGQENKGDVGPRTSQ0GSP-PPVN 1630  
 QY 907 NSS-----PYSVIPQPGM-----MGNGMGIG-----NQG 931  
 DB 1631 LSAVAGPPPSLP--GLIGLSALGAGIGGPGKGMHFYFNFIPPAYVYNDVPRGS 1689  
 QY 932 LGNSTGTMGINSASRPMPSEEMAPQSSAVAYVTCATTSAMNRPVQGMIRNPAASIPMR 991  
 DB 1690 VSIYASEEAAKLGGHPGLPPSSQAOLSGISIK-----ERLKESSP--SPHQ 1735  
 QY 992 PSSQPGROTLOSQVNNIGP--SELEMMGGQ0Y0Q0QAPNQTAPWPESTLPIQ0AFAS 1050  
 DB 1736 PKHNPSSQ0MIAKLIQ0EPMTK0EIK0EPNSNPQ00HPR00Q0PAP0--PQ00Q0PPP 1792  
 QY 1051 QNRQPFSSPDDLCLCPRAAESPSDEGALLDQLYALRNFDGLBEIDRALGIPELVQS0Q 1110  
 DB 1793 QPQ0PHALHPKDL-----QALGAYPAIYQRH 1818  
 QY 1111 AVDPQESSQDSNIMLEKQAPVFPQ0Y-----ASQ0MAQ0SGSYTMQDPNPHM 1159

DB 1819 ST-----NLAVQ0AREEELRRYMTGTRONSAAAAA0A0AAS----- 1858  
 QY 1160 GORPSYATLMQRRPGLRPTGLVQONP0DLQLD-QHRLQ0-Q0N0PLMNQISNVSNV 1217  
 DB 1859 GGLRPHGMMHKREPGSGA0Q0Q0Q0Q0Q0MDLQ0Q0Q0A1Q0H0H0LQ0Q----- 1910  
 QY 1218 LTLRPGVPTQAPLNAQMLA0REILN0LRC0RM0Q0Q0V0QRTLMKQ0----- 1268  
 DB 1911 -----HQADQ0Q0Q0Q0H0Q0Q0Q0Q0Q0Q0Q0Q0Q0Q0K0Q0A0SAGANNK 1958  
 QY 1269 GLNMT-----PSVAPSGMPATMSNPRIPQANAQ0FPPPNVGISQ0PDP-G 1314  
 DB 1959 ATNLTQDSPKQGGDDQPLKQ0EGKPTMEQ0GPPRPTSQFLHPST-IS--P1P1G 2015  
 QY 1315 FTGATTPQSPILMSPRMHTQSPM0Q0QANP---AYQAPSDINGMAQGNMGNSM----- 1366  
 DB 2016 F-----DPNHMYRNVNLSAAGPYNTAPYHLPIRPHNAPRDLs-----RNTGTALADLH 2067  
 QY 1367 -FSQ0QSPH-----FGQANTSMYSNNMNIIVSNAT-----NTGMSNNQMTQDISMYS 1415  
 DB 2068 AASQYTTTHKIHLSERALKSPTSQGPVKVSVSPSIGPPQGGPPTSGSGSPVS--- 2124  
 QY 1416 VTSVTSGLSSMGPEQVNDPALRGNI-FPNOLPG 1449  
 DB 2125 -----GVLDGPGS-----GSMQ0PQSGAPG 2142

## RESULT 30

Q9WVS9 PRELIMINARY; PRT; 862 AA.  
 AC Q9WVS9;  
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
 DE CLOCK.  
 GN Rattus norvegicus (Rat).  
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxId=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=99197211; PubMed=10095082;  
 RA Abe H., Honma S., Namihira M., Tanahashi Y., Ikeda M., Yu W.,  
 Honma K.;  
 RT "Phase-dependent induction by light of rat Clock gene expression in  
 the suprachiasmatic nucleus";  
 RL Brain Res. Mol. Brain Res. 66:104-110(1999).  
 DR EMBL; AB019258; BAA81819.1; -;  
 DR HSSP; P36956; IAM9.  
 DR InterPro; IPR001092; HLH dim.  
 DR InterPro; IPR003015; HLH Myc.  
 DR InterPro; IPR001067; NucleusLocator.  
 DR InterPro; IPR001610; PAC.  
 DR Pfam; PF00785; PAC; 1.  
 DR Pfam; PF00989; PAC; 2.  
 DR PRINTS; PR00785; NCTRSL0CATR.  
 DR SMART; SM00353; HLH; 1.  
 DR SMART; SM00086; PAC; 1.  
 DR SMART; SM00091; PAS; 2.  
 DR PROSITE; PS00038; HELIX\_LOOP\_HELIX; UNKNOWN\_1.  
 SQ SEQUENCE 862 AA; 97003 MW; D031EA43758907EC CRC64;

Query Match 4.3%; Score 328; DB 11; Length 862;  
 Best local similarity 21.0%; Pred. No. 1.1e-11;  
 Matches 200; Conservative 149; Mismatches 370; Indels 232; Gaps 43;

QY 31 KRTEKRNROENKYEIELLELIFANFNDINFNKDKCAILKEVYKQIRQIKQEKKA 90  
 DB 39 RNSEKRRQ0FNVLLIKELGSLPGRNAR-----KMDKSTVLQKSIDFLR--KHKEITA 89

```

0Y 91 AANIDEVQKSDVSSYSGOVIDIKDALGRMILALDGFVVALBENNVFVSEBNTQYLRYN 150
Db 90 QGDASEI--RODMKRP---FLSNEEFTQMLBEGFLALMTDSDIIVSEVYLSLEHL 145
0Y 151 OELMANKSVYSLIHHGDHTEFEVKNNLPKSIYNGSGWSGEPPR--NSHSTFNCRMJ----- 203
Db 146 PBDVLDQSIFFNPIRGEHSEYVK--ILSHLLESDSLTPEDLKSNOLEFCCNHLRGITDP 204
0Y 204 -----VKLPDSEEBCHDN--OEAHQ-KYETMOCPA---VSOPKSIK 239
Db 205 KEPSTYEYVREIGNFKSLNSYSTSHNGFEETIORTHRPSYEDHVCVATVRLATPOFIK 264
0Y 240 EEGEDLOSLCICVARARVPMKREPYLSESTETRODLOGKTTSLDSTMTAANKPGMEDL 299
Db 265 E-----MCTV-----EER-----NEETSHSLKMKFLYD---HRAPIIGLPP 302
0Y 300 VARRCIQTHAOGHEGESVYAKRHHEVYLRQSLAFQYRFLSLSDGLVYAQTKSKLIRSQ 359
Db 303 EYLGTSYGDYVHVLDLESLACKCHN--LMQYKGSCKSYRFLTKQOQWIMLQTHYIITYHO 361
0Y 360 TTNEPOLYSLIHLML-----HREONCVMMPDLTGOTMGRPLNPISSNPAIDALC 409
Db 362 WNSRPEFIVCTHYVYSTA EVAERRELRGVEESLPE-----TPAAKSOD 405
0Y 410 SGNPQODMTLLSNINFRIPNGPEOMGMKRFSGSGGMNHYSGMQATTPOGSNYALAKMS 469
Db 406 GSGDNRIYTVS-----LKLEALREFDHS-----PTPBASSRSRKS 441
0Y 470 -PSQSSPPAMNQG--PTMSLSPRHMKSPCVAGSPIPPSQSPAGSLHSPVGCSSTGNSH 527
Db 442 HTVAVDSPSTPTKLPDTDSTRPR-----PILPAPHE-----KMTORRS 478
0Y 528 SYTNSLLMALOALBEGHGVSLGSSLASPDLMKMGILQMSPPVMMNPPLS--KMGSLDS--KDC 585
Db 479 SFSSQSINS-----QSVGSSLLTPRMSQAMNLPITPOMKQFOLSAQLGAMQHLKD-- 528
0Y 566 FGLIGPESEGITGAESSCHAPGEDEKFTNDPMLPRAVSSERADG--QSHLDSKQTKL--L 642
Db 529 -----OLEORTMEIANIH--ROOEELR--KIOEOLQMVAGOGLOMTLOOSNFGMLNGSV 579
0Y 643 QLTFTKSD--OMEPSPLASSLSDNNKSKTSLSPSGSTHGTSLKEKHKLILHRLDSSSP 700
Db 560 QLSGNSNIQDLITINNOGOVVPVNOIQSGVGNAGHST-----GOMHIOOQTLQSTSTQ 633
0Y 701 VDLAKLTAELATGKDLSESSS--TAGSEVYTKOEVPSPKKENALLRYLDDKDTKDIG 758
Db 634 SQQSGVMSGHSPQLSPNOTESTPLAPLYNTMVISOPAGSM-----VP 676
0Y 759 LPEITPKLERDSTDPDASNTKMLAMKTEKEMSFERGQDPGSELDNLE-----EILD 812
Db 677 IPSSMPO-----NSTQSAVYVTF--TORQIRFSGOOLVYKLVATAPVAGAVVPST 727
0Y 813 LQNSQLPOLPFP---DTRPGAPAGSVDKOAILINDLMQTLAENSVPVTVGAOKTLARISOST 869
Db 728 MLMGOVYATVYTFATQOQQAALSVTQOQ--QOQOQOQOQOQOQOQOQOQOQOQ--QSQQSS 783
0Y 870 FNNRRPG-----QLGRLL--PNQNLPLDITIQSPFGAGPEPP 904
Db 784 QDQHPHPSVQOAPOLTRPOOPFLQTSRIILHNNPSTOLLSAAPFLOQSTPP 834

```

RESULT	31	
Q91YB0		
ID	Q91YB0	PRELIMINARY;
AC	Q91YB0	PRT: 865 AA.
DT	01-DEC-2001 (TrEMBLrel. 19, Created)	
DT	01-DEC-2001 (TrEMBLrel. 19, Last sequence update)	
DT	01-DEC-2001 (TrEMBLrel. 19, Last annotation update)	
DE	CLOCK PROTEIN.	
GN	CLOCK.	
OS	Splax gallii.	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Spalacinae;  
 OC Spalax.  
 OX NCBI\_TaxID=164333;  
 RN 11  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=BRAIN;  
 RA Avivi A.;  
 RT "Biological Clock in Total Darkness:The Clock/MOP3 circadian system of  
 the blind subterranean mole rat.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 0:0-0(0).  
 SO EMBL; AJ318057; CAC85403.1; -.  
 SR SEQUENCE 865 AA; 97437 MW; E5003919187578C21 CRC64;

Query Match	4.3%;	Score 328;	DB 11;	Length 865;
Best Local Similarity	20.0%;	Pred. No. 1.1e-11;		
Matches 220;	Conservative 160;	Mismatches 370;	Indels 350;	Gaps 53;

QY	31	KRNEKRNREQNKXKIEELAEELIFANFNPDIDFNKRPCKCALKEVAKOIROIKEDOKA	90
Db	39	RKSKSEKRRDKDQFNVLKEIGSMLPGNAR-----KMKSTYLOKSIDELR--KHKETIA	89
QY	91	AANIDEVOKSDVSSGQGVIDKALGPMMLLEADGFFEVFNLEGNVFESEVANTOYLKRN	150
Db	90	QSDASEI--RQMKPR---FLSNEETQMLELDDFFLAIMDGSIITYSEVSTILEHL	145
QY	151	QEELMKNSVYSLIHVGDHTEFYKNLLPKPSIYNGSGWSGEPPR--RNSHFNCRML-----	203
Db	146	PSDLDOGVFNFIPEGEHSEYVK--ILSTHILLESDSLTPREYLSKKNQLEPCCHMLGTVDP	204
QY	204	-----VKPLDSEBEGHDN--QGAHQ--KYTMOCEFA-----VSQPKSLK	239
Db	205	KEPSTEYEVRFIGNFKSLNSVPTSAHNCFEGTQIRHRSYEDRVCFAVATVRLAPPOFK	264
QY	240	EEGEDLOSLICVARRVPMKERYPVSSESEFTRODLOGKITSLDTPSRAMAKGWMEL	299
Db	265	E-----MCTV-----EER-----NDEFTSRISLEMKFLFD--HRAPITIGYLPF	302
QY	300	VRCIOKTHAOHEGESVYAKRHHHEVLROGLAFSQTIFYSLSDGTLYAAQTKSLIRSQ	359
Db	303	EVLGTSGYDYHVVDLEMLACKEH--LMQYCKGKSCYYRFLTKGQGMWIMQTHYIITYHQ	361
QY	360	TTNFOLVLSLHML-----HREQNVCMNBDLQGTMRKRLINISSNPRHQAOLC	409
Db	362	WNSRPERFVCGHITYAVIEVRAERREGLIEBSLPD-----AAADKQSD	405
QY	410	SGNPGQDWTLLSNINFPINGPKQEMGMGRFGSGGNHVSQMATTPOGSNYALKNMS	469
Db	406	SGSDNRITVS-----LKEALERFDH-----PPPSASSRSKRSS	441
QY	470	-PSQSSPGCMNPGQ--PTSMLSPRHRMSPGVAGSPRIIPSOFSAGLSHPVGCSSGTGNSH	527
Db	442	HTAASDPSTPTKIPITDST-----PRQHLRA--HEKMAORRS-----	478
QY	528	SYTNSLSALALABGHVSLGSSLASADLKMGNLONSPYVMNPPRLS-----KMSLD	581
Db	479	SFSSQSMNS-----QSVGPSTLQD--VMSQAALNPV---PQGMQFOFSAQIGAMQ	524
QY	582	S-KDQFCGLYGESEGTGOAESCHPGEOKETNPDLNPRAVSEERADQ--QSRILHOSK---	636
Db	525	HLKD-----QLEORTMIEANIH--RQEBELR--KIOEOLOMHNQSGLOMFLOOSNGCL	574
QY	637	--GQTKILLQTLTTSKDOMEPSPLASSLSDTINKDSTGLPGSGSTH--GTSLEKHKILHRL	693
Db	575	NFGSVQLSSGNSNIOQLTPIIMQGOVYPTNQIOGSMNAG---HIGHS---QHLIQOQS	627
QY	694	LQDSSSPVDLAKLAELATGKLDSESSSTAGSEYTTIKQEFVSRKKENALLRYLLDKD	753
Db	628	LQSTS-----PQOSQOVSVMGSHOOTSLASQTO-----	655
QY	754	TKDIGLEPITPKLERLDSKTDPRASNTKLIAKTEKEKSFEPGPOGSELDNLEIIDL	813
Db	656	-----SITABUYNTAVIS-----QRA--PQSMV-----	672



```

OY 1192 -----OLOHRLQAOQNROPPLANOISNVSNVNLTLRPGVTPQAPINAQMLAORREIL 1243
DB 3890 LHAQOHPTQKHQAQOQOONQOIQOHQSOQ-----QHVOQOQNAQOQHLSQ 3935
OY 1244 NOHLRQRMHQOQOYQOQTL-----MARG-QGLAMTPSMVAPSG-----MPATMSNR 1290
DB 3936 OOHQSOOOLNOHQOQOQOOLQOIQLOQMHGPOOQKSPQGVGHGSGTSLFASOQHSNQ 3995
OY 1291 IP-----QANAOQF-----PPPPNYGIS-----QOPDPGFGTATTPQSPMLSPRMAHTOS 1335
DB 3996 LPARVPOOQHPOQLSHSPCKPPTLVSNOCVOPPALITRVGSHSOPRQOQOQLPHQOSS 4055
OY 1336 ---PMAQSO---ANPAYOAPSDINGMAQNGMGSMSFQSPPHFGOQANTSMYSNNMN 1389
DB 4056 SGHPHQKQLSSPGANLPLPTPLNVIONTPKTIYQOHIVAQNVPPPOQVNAIHTPONG 4115
OY 1390 INVSATVTGMSNMONTGQISMTSY-----TSVSTGLSSGPPQVNDPALRGNLPPN 1445
DB 4116 KD---STPPGHVEPTPPAMSAQKTSSESVIRTPPTTGLAVISANTVGS-LITTEENLKI 4171
OY 1446 QLPQMD-MIKOE 1456
DB 4172 SOPKODELIBD 4183

```

```

RESULT 33
O91YB2 PRELIMINARY; PRT; 865 AA.
AC O91YB2;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE CLOCK PROTEIN.
CN CLOCK.
OS Spalax carmeli.
OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Spalacinae;
OC Spalax.
OX NCBI_TaxID=164324;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RA Avivi A.;
RT "Biological Clock in Total Darkness:The Clock/MOP3 circadian system of
RT the blind subterranean mole rat.";
RL Proc. Natl. Acad. Sci. U.S.A. 0:0-0(0).
DR EMBL; AJ318058; CAC85404.1;-.
SQ SEQUENCE 865 AA; 97419 MW; CA58A56F535C6625 CRC64;

```

Query Match 4.38; Score 325; DB 11; Length 865;  
 Best Local Similarity 19.98; Pred. No. 1.7e-11;  
 Matches 219; Conservative 161; Mismatches 370; Indels 350; Gaps 53;

```

OY 31 KRNRKRRREOENKYEELAEILFANFNIDFNFKPKCALIKETVQKIROIKFOEKAA 90
DB 39 RNKSEKRRDDQFNVLIRKLSGMLPGNAR-----KMDSTYLOKSIDPLR-KIKETITA 89
OY 91 AANIDEVOKSVSTGCGVLDKDALGPMLEALDGFVVNLNENGVNVSENVTOYLRYN 150
DB 90 QSDASEI-RODMKPT---FLSNEEFTQMLEALDGFPLAIMDGSIIYVSESVTLLEHL 145
OY 151 QEELMNKSVSILVGHDETEYVKNLPRKSIYNGSGMSEPPR-RNSHFNCML----- 203
DB 146 PSDLVDSVFNFIPGEHSEYK-ILSTHLESDSLTPRYLKSQKQLEFCCHMLGCTIDP 204
OY 204 -----VKPLPDESEEGHDN--QEAHO-KYETMQCFA-----VSQPKSIX 239
DB 205 KEPSTYEVVRFGNFKSLNSVPTSAHNGFEGTIOGTHRPSTVEDRCVATVYLAIPQIK 264
OY 240 EDEGDIQSLCIVARRVPMKEBPVLPSSSEFTTRDLOGKITSLDTSTWRAAMKQWEDL 299
DB 265 E-----MCTV-----EPP-----NEEFTSRHSLMKFLFLD---HRAPIIGYLPF 302

```

```

OY 300 VRCIOKFAHQEGESVSAYAKRRHHHEVLROGLAFSOIYREFSLDGTLVAAOTKSLIRSO 359
DB 303 EVLGTSGDYIYVDDLENLAKCHEH-LMQYGGKGCYTRFLTKGQWIMLQTHYIYTHQ 361
OY 360 TTNEPOLYISLML-----HREQNVCMVNDLTGQTMGKPLNPLISSNSPAHQALC 409
DB 362 WNSRPEFICTHTVSYAEVRAERRELIGIESLDP-----AAADQSD 405
OY 410 SCNPQODMTLSSNINFPINGKEQMGMPRGKGGSGGMNHSGMQATTPOGNTYALKNS 469
DB 406 SCSDNRINTVS-----LKEALERDHS-----PPSASSRSRKS 441
OY 470 -PSOSSPGMNGQ-PTSMLSPRHRMSPVAGSPRIPPSQFSGLSHSPVGCSTGNSH 527
DB 442 HTAVSDPSTPTKIPITDST-----PPROHLPA---HEMMAORS----- 478
OY 528 SYTNSLNALQALSEGHSGLSSLASPDLMKGNLQNSPVNNPPPLS-----KMSLD 581
DB 479 SFSQSGMNS-----QSVGPSLTQPIYISQA--ANLPV---PGMGQFQPSAQLAGAQ 524
OY 582 S-KDQFGLYGEPSSEGTGOAESCHPECKETNDNLPVAVSSERADG-QSRHLDK--- 636
DB 525 HLKD-----OLEORTRMIEANIH-RQEEELR--KIOEQLQMVHGQGLQMFLOQSNPGL 574
OY 637 --GQTKLQQLLTTKSDOMEPSPLASSLSDTNKNDTSGSLPGSGSTH-GTSLKKEKHLRL 693
DB 575 NFGVQLSSGNSNIQQLTPIPMGQVPTNOIQSGMNG---HIGTS---QHLIQOOS 627
OY 694 IQDSSSPVDLAKLTAEATGKDLQSSSTAPGSEVYTIKQEPVSPKKENALLRYLLDKD 753
DB 628 IQSTS-----TQSSQSVMSGHSQOTSLSAQO--- 655
OY 754 TKDIGLPEITPKLERLDSKTDPASTKLIAMKTEKEMSFEFGDPGSELDNLLEILDL 813
DB 656 -----STLTAPLVNTWYIS-----QPA--PGSVV----- 677
OY 814 QNSQLPQLEPOTRRPAPAGSVQKAIINDMLQJTAENSVPTVYGAQKATLRISOSTFNNP 873
DB 678 --QIPSSMP-----QNSTQSAVITTF--QODROIRSO----- 706
OY 874 RPOQLGRLLPNQNLPLDITLQSPGTAGPFPPIRNSSPYSVIRPQMGMGNGMIGNGLG 933
DB 707 -----GQQLVTK-----LVTPAPVACAV-----MVSITLMQV----- 735
OY 934 NSTGMIGNSASRPTMPGEGMAPQSSAVRYTCATTSAMNRPVQGMIRNPAASIPMRPS 993
DB 736 -----VTAYPTFATQOQQAQTLV-----TQOQPOQ---QOPOQOQPOQOQ 773
OY 994 SQPGOROTLQSGYVNI-GPSELEMMMGCPQYSSQQA-----PPNGTAPWPELIRIDAS 1047
DB 774 PQOQOQSSQOEQQLPSVPOPQOQLQSPQOQLQTSRLLHGNSTOLI--LSAAPLQOST 831
OY 1048 FASQNRQPFSSPDDLLCPH 1067
DB 832 PPSHHQOQSOQOQOQLSRH 851

```

```

RESULT 34
O91YAB PRELIMINARY; PRT; 865 AA.
AC O91YAB;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE CLOCK PROTEIN.
CN CLOCK.
OS Spalax judaei.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Spalacinae;
OC Spalax.
OX NCBI_TaxID=134510;
RN [1]

```







```

OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=RETINA;
RX MEDLINE=99337784; PubMed=10407173;
RA Larkin P., Baehr W., Semple-Powland S.L.;
RT "Circadian regulation of iodopsin and clock is altered in the retinal
RT degeneration chicken retina.";
RL Brain Res. Mol. Brain Res. 70:253-263(1999).
DR EMBL; AF132531; AAD43283.1; -.
DR HSSP; P36956; 1AM9.
DR InterPro; IPR001092; HLH_dim.
DR InterPro; IPR003015; HLH_Myc.
DR InterPro; IPR001067; NucleusLocator.
DR InterPro; IPR001610; PAC.
DR InterPro; IPR000014; PAC.
DR Pfam; PF00785; PAC; 1.
DR Pfam; PF00989; PAS; 2.
DR PRINTS; PR00785; NCTRNSLOCATR.
DR SMART; SM00353; HLH; 1.
DR SMART; SM00086; PAC; 1.
DR SMART; SM00091; PAS; 2.
DR PROSITE; PS00038; HELIX_LOOP_HELIX; UNKNOWN_1.
SQ SEQUENCE 875 AA; 98725 MW; 04DFEB1D79747A4 CRC64;

```

Query Match	4.28;	Score 319;	DB 13;	Length 875;
Best Local Similarity	19.28;	Pred. No. 4.1e-11;		
Matches 189;	Conservative 137;	Mismatches 363;	Indels 294;	Gaps 37

OY		31	KRNPEARNREONKYEIELAEELIAFNFDNDNFNPKDCKALKEVTKOJIKEDSKAA	90
Db		39	RKSEKKRRDDPNVLKELGSMLEGNAR-----KDKSTVKLSIDFLK--KHKEITA	89
OY		91	AANIDEVOKSVSSSTGGGVINDKDALGPMLIEBALDGFEVVUULEQNVFSENVTQYLRYN	156
Db		90	OSDASEI-RQMKPT---FPNEBEFTQMLEADGFPLAINTDSNIITYBSVTPPLEHL	145
OY		151	QEELMKNKSIVSILHVGDTTEFFVKMLLPKRSIVNGSGSWGEPPR-RNSHTPNCRL-----	203
Db		146	PSDLVDQSVEFNFEDEGHSEIYK-ILSSHLLIESDLPREYKLKSNOLFCFCHMYRTIDP	204
OY		204	-----VKLRDSEEBEHDN--QEAHQ-KETMQCF-----AVSQPRSK	239
Db		205	KEOPTYEYVKFIENFKCLNNPVNAHNNEBETIDRSHSPSEDKVCCATAYLVALTPOPIK	264
OY		240	BEGEDLOSCLICVARRVPMKERPYLPSESSEFTTRDODLOGKITSLDTSIRAMAKPWEDL	299
Db		265	E-----MCTV-----DEP-----NEBTSHSLMEKWLFLPD-----HARPILIGLPF	302
OY		300	VRCIOKTHAOHEGESVYAKRHNHVELRGSLAFSOIYRESLSOGTLVAATSKLIRSQ	355
Db		303	EVLGTSGDYHHVVDLDIMLACKEHE-LMOYKGKGSICYAREFLTKGOOMIWIOTHYITYHQ	361
OY		360	TTNESQULVISLHML-----HRBNQCVMMPDLTGTMGRPLNPISNSPAHALC	409
Db		362	MNSREFVTCVTHTVSYAVEBRAERRRELIGIESLEPI-----	398
OY		410	SQNPQODMTLSSINIFPINGRKEQMGMPRGFGSGGMHVSGMOA-----TTPQGS	461
Db		399	KADKQD-----SGSDNHINTVYSLKALELERPDTSP7PS	431
OY		462	NIALKMNSPQSSPGCMNFGQPTSMLSPRHMSPGVAGSPRI7PSQFS7PAGSLHS7PVCVS	521
Db		432	ASSRSSRKRKSHTA5VADHDSSTPTKM-----TYDTPRPQO--SLSAHE-----K	472
OY		522	STGNHSTJTNSSLNALQALSECHEGYSSGLSSLASDPLKMGNIQNOSPVMNPPPL--SKMG	578
Db		473	STORRSJLSSOSLSS-----OSTLOPOTOPMSOPATIOHSSSQVFOFQASOLD	523

```

0Y 579 SLDS -KODFGXGPERSEGTGQAASSCHPBGOK----- 610
Db 524 AMOHLKD-----QLEORTMTAMANIHROOEERKIQEOLQIYHGGGLDSSGPCVPKI 576
0Y 611 ETNPENLPAVSSERADQO--SRLHDSKGOTKILLOLTTKSDQMEPSPPLASSLSDTNKDS 668
Db 577 HRTDVTVEMLQOSTSGLNSSFVQLTSGNSSVQ-----QLAPENMGQVVOVINQO 629
0Y 669 TGLSPGSGSTHGTSLKEKHKTILHRLLODSSSPVDLAK-LTAEATGKDLSESSST--APG 725
Db 630 SGMNTHST-----PHMIQQPLPSSASQHNQAVLSGHGQSSLAGOSONFTVSTPL 682
0Y 726 SEVITKOEPRSPKKENALLAYLLKODTGDIGLPEITPLERLDSKTDPASTYKLIAM 785
Db 683 YNTMYISOPTGANN-----VOYPSLPO-----NNQNAAAATTF 717
0Y 786 TEKEMSEFPDOPGSELNLEIILLDQ-----NSQLPOLPEDR 826
Db 718 TQDRKIRRSQGOQLYTKLYTPAIVACGAVMPSYTFMGQVYATPTFAAQOQOQOTLEPVQ 777
0Y 827 PGAPAGSVDKQAIIINDLMQLTAENSPLYTAVGAQ-----KTALR 864
Db 778 QQQQQQQQSQDDQQQQQQOLTAFAVQAPAPOLQTHPQOLFQTSRLHGNQSAOLILSAFP 837
0Y 865 ISQSTF-----NNPRGQGR 880
Db 838 LQOSTFTQSHHQOHOPQOQOLSR 860

```

RESULT	38
Q9W6Q2	
ID	Q9W6Q2
PRELIMINARY;	
PRF;	853 AA.

```

DT      01-NOV-1999 (TREMBLrel. 12, Created)
DT      01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT      01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE      CLOCK PROTEIN.
OS      Gallus gallus (Chicken).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Actinopterygii; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC      Gallus.
OX      NCBI_TaxID=9031;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=20490740; Pubmed=10931848;
RA      "Chong N.W., Bernard M., Klein D.C.;
RT      "Characterization of the chicken serotonin N-acetyltransferase gene.
RT      Activation via clock gene heterodimer/E box interaction."
RL      J. Biol. Chem. 275:32991-32998(2000).
DR      EMBL; AF144425; AAD32860.1; -.
DR      HSSP; P36956; 1AM9.
DR      InterPro: IPR001092; HLH_dim.
DR      InterPro: IPR003015; HLH_Myc.
DR      InterPro: IPR001067; NucleinsLocator.
DR      InterPro: IPR001610; PAC.
DR      InterPro: IPR000014; PAS.
DR      Pfam; PF00785; PAC; 1.
DR      Pfam; PF00989; PAS; 2.
DR      PRINTS; PR00785; NCRNSLOCATR.
DR      SMART; SM00353; HLH; 1.
DR      SMART; SM00086; PAC; 1.
DR      SMART; SM00091; PAS; 2.
DR      PROSITE; PS00036; HELIX_LOOP_HELIX; UNKNOWN_1.
SQ      SEQUENCE   853 AA;  96297 MW;  534965C1F7293C97 CRC64;

```

```

Query Match      4.2%  Score 318;  DB 13;  Length 853;
Best Local Similarity 21.7%  Pred. No. 4.5e-11;
Matches 165;  Conservative 116;  Mismatches 249;  Indels 232;  Gaps 37;

QY  31 KNTKERNRQEGNKIYEELAEELIFANFNNDINDNFRPKDCAIKETVQVRIQIKDERAA 90
      : 111: 1:1 1:11 :: 1 11 :111: :1 1:11 :

```

Db	39	RKSSKKRRDQPNVLIKELGSMRLGMA-N-----KMDKSVTLQKSIDFLR--KHKEITA	89
QY	91	AANDVQKSDVSSSTGQGVIDKDALGRMMLDALDGEFFVNVLEGNVYVSENTOYLRYN	156
Db	90	OSDASEI--ROQMKRP---FLSNEEFTQIMLEADLGGFALMTDGNITIVYSESVPLLEHL	145
QY	151	QELMKNKSVYSILVHGDTTEFEVKNLLPRPSYVNGGSMGSEPPR--RNSHPNCRMI-----	203
Db	146	PSDLVDQSVFNFTIQGENSEIYK--ILSHLLESDSLPRPEYIKSKNNQLEFCCHMLRGITDP	204
QY	204	-----VKRLPDEEEEDHON--OEANO-KYETMQCF-----AVSQPSIK	239
Db	205	KEQPTVEYKFLGNFKCILNNVPNSAHNGFECTIQSRHPSYEDKVCFLATVRLATPQFIK	264
QY	240	EEGEDLOGLCVARARVPMKRRPRLPSESESTTQODQKLTISLDTSTMAAKMPGEDL	299
Db	265	E-----MCIV-----EER-----NEETSHTSLMEKFLFID---HRRPPIIGLPR	302
QY	300	VRRCIOKFEHAESESVYAKRHHNEVLROGLASFQYREFSLDGTTLVAQTKSLIRSQ	359
Db	303	EVLGSGDYHHVDLDLNLACHEN--LMQYKKGKSCYYRFLTKQOQIMTLQTHYITLHQ	361
QY	360	TTNEPOLYISLHML-----HREDONCVAMPDL---TGQTMGRP--LNP1-----	398
Db	362	WNSRPEEFTYHTGTAVSYAEVRAERRRRELGEESLPEIKADKSODSGSDNHIINTSLKEAL	421
QY	399	-----SSNPAAHQL-----CSGNPGQDML-----	419
Db	422	ERFDTSPRPSASSSSSKRSSHITAVSDHSSITPTKATVTSTPRRPSLHNEKSTORSSLS	481
QY	420	-----SSNINPRINGPKR-----QMGMPMGRFGGS--GGMNHV-----	450
Db	482	SQSLSSGSLQPVYQPTMGSPQATLQLOSSMGOFSAGQAGMCHLKQLEQRTBMEANIR	541
QY	451	-----SGMATTPQ---GSNT--ALMKNPSQGS-----PGANPGQPTS	484
Db	542	QOELRKIQEOLQIVHGGLQMLFLOQSTSGINFSSVOLDTSGNSSSYOOLARPGNMGGYVO	601
QY	485	MLSPRHKMSPPVACSPRIIPRQSPAGSLHSBPGVSCSTGSHSYTNSLNALQALSEGH	544
Db	602	TNQTQSGKNTHGTSPTPHIQOQ-----PL---QSSASQHNQON-----VLSGH	641
QY	545	GVSIGSSSLASDCLKMGNLONSPPVMMPPPLSKMGSLDKDFGLYGPSEGTQGAESSC	604
Db	642	GQO---SSSLA-----GOSQON---VSTPLVNTM-----VISQPTGNVQVPSSTL	680
QY	605	HPGEQKETNDENLPPAVSEKADQSGSRHLHOSKQGTOKLLOLT	646
Db	681	-----PQNNNON-AAATVTTFTQDKQIRF---SQGQOLVTKLVT	714
RESULT	39		
Q9XTA4	Q9XTA4	PRELIMINARY;	PRT; 870 AA.
OC	09XTA4		
OC	09XTA4		
OC	01-NOV-1999	(TtEMBLrel. 12, Created)	
OC	01-NOV-1999	(TtEMBLrel. 12, Last sequence update)	
OC	01-DEC-2001	(TtEMBLrel. 19, Last annotation update)	
OC	ENDOTHELIAL PAS DOMAIN PROTEIN 1/HYPOXIA-INDUCIBLE FACTOR-2		
OC	ALPHA.		
OC	EPAS1/HIF2 ALPHA.		
OC	Bos taurus (Bovine).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eulhelia; Cetartiodactyla; Ruminantia; Pecora; Bovidae;		
OC	Bovidae; Bovine; Bos.		
OC	NCBI_TaxID=9913;		
OC	[1]		
OC	SEQUENCE FROM N.A.		
OC	TISUB=ARTERY;		
OC	MEDLINE=99255430; PubMed=10320777;		
OC	Hara S., Kobayashi C., Imura N.;		
OC	"Molecular cloning of cDNAs encoding hypoxia-inducible factor (HIF)-		
OC	1alpha and -2alpha of bovine arterial endothelial cells."		

```

QY 725 -GSEVITIKQ-----EPVS-----PKKENALLRYLLDKDDT 754
DB 610 GGSRSYSLQCCGQTYPTPLSMGSGISNTQWPPDPLQLGPTKMGCEDRHA-----SAVGA 653
QY 755 KDIQLEIRPKLERLDSKT-----DPA-----SNTKLIAMKTEKEMSFEP--- 795
DB 664 APLGLPPLPHTLMLKKRSKSGFGPGCPDYMSAMALSKLKLKQLEEDQAFQDMSG 723
QY 796 GPGPSELNLE-ELIDDLQNSQLPOLFPDTRPGAPAGVDAQIINDLMQTLAENSPLY 854
DB 724 GPPGSGTSHLMWKRMSKSLRGSGTSLMPDK----- 754
QY 855 PVGAQALRISOSTFNNRPPGOLGRLLPNQNLPLDITLQSPGAGPPPIRNSPSYVI 914
DB 755 -----LNNANVPNDEFIONPV-RGRSOPFLRHLSP----- 782
QY 915 POPGMNGMIGNOGNLGSSGTGMSASRPT---MPSGEWAPQSSANRYVCATTS 971
DB 783 POP-----PSAT-----SPEPTKSGFPQCYAPOTDYSLPAHKMSG 821
QY 972 MNRPVGGMIRNPASIPMRSSOP 996
DB 822 M-----ASRLLGPFEP 833

```

```

RESULT 40
QVEG7 ID QVEG7 PRELIMINARY: PRT: 2703 AA.
AC 09VEG7
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE OSA PROTEIN.
GN OSA OR CG7467.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota: Metazoa: Arthropoda: Tracheata: Hexapoda: Insecta:
OC Pterygota: Neoptera: Endopterygota: Diptera: Brachycera: Muscomorpha:
OC Ephydriidae: Drosophilidae: Drosophila.
OX NCBI_TaxId=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RA MEDLINE-20196006; PubMed-10731132;
RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amaralides P.G., Scheer S.E., Li P.W., Hoskins R., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Morten J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borotova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Daveyport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durkin K.L., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Houslin D., Houslon K.A., Howland T.J., Wei M.-H., Ibeagwu C.,
RA Jalili M., Kalush F., Kapen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laseo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Mervulov G., Mlshina N.V., Modyarty C., Morris J., Mostrel A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Murzy D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Relbert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,

```

```

RA Svitskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodade T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL: AE003718; AAF55457.1; -.
DR FlyBase: FBgn0003013; osa.
DR Interpro: IPR001606; ARID.
DR Interpro: IPR002965; P_rich_extensn.
DR Pfam: PF01388; ARID.1.
DR PRINTS: PR01217; PRICHTENSN.
DR SMART: SM00501; BRIGHT.1.
SQ SEQUENCE 2703 AA; 282894 MW; 27ED5BA0277A6369 CRC64;

```

```

Query Match 4.1%; Score 314.5; DB 5; Length 2703;
Best Local Similarity 20.1%; Pred. No. 3,9e-10;
Matches 245; Conservative 111; Mismatches 475; Indels 387; Gaps 48;

```

```

QY 392 GRPLNPIS-----SNSPAHQALCSGNPGODMLSSNINFPIN--GPKEGMGMPMRFG 443
DB 676 GPPRPPTSOAGACGAGSMPSGAQAGTGGKMPNHTGQIPRYQWVPSPQOTVPGAGPG 735
QY 444 SCGMNHVSGMQA-----TPPGSNYALKNMSPSQ-----SSP----- 475
DB 736 AMGNHIVGCGTPRPVYGVGPPRPGSGSPRLNYLKHQHGKGVGSGTPPGQGYG 795
QY 476 ----GMPGPTISMLSPRHRMSPGVAGSPRTPPSQSPAGSL--HSPYGVCSSTGNSH 528
DB 796 NGPTGMNPGMP--MGRPHNMGPRHGTNNMGPPTSTPSPQSLQGGQPOGQASG-- 848
QY 529 YTNSSLNALQALSEGGVSLG--SLASPLDKMGNLQNSPVNNPPLSKGSLDKCF 586
DB 849 ---PESGGRHISQDNGISSGPTGAAGHAYTSVTTGPDGTSMDVSSQSTLSNMSAA 905
QY 587 GLYGPSEGTGQAES-----CHPG-----DQKETNDNLPYAVS 622
DB 906 S--GDPQCTTPRSKRNDPYSQSHLAPRSTSPRVYMHGCGEGEYDMSSPPNWRPAG 963
QY 623 SERADGOSRLHDSKQGTLLQLLT--KSDMPERPLASSLSTPNKDSGLSGSGSTHG 680
DB 964 SPQVFN---HVPVQPEPFRSTYTTTKKSDSL-----CKLYEMDNDPRRGMW----DKLR 1011
QY 681 TSLKEKHILRLDQSSSPVDLAKL-----TAATGKDL-----QESSST 722
DB 1012 AFMEERKPTTACPTTSKQPLDLYRLIYKERGFEVETKSTWKDIAGLIGASSSA 1071
QY 723 APGSEVITIKQEPVSPKKNALLRYLLDKDDT-----DIGLEITPKLERLDS-- 771
DB 1072 A-----YTLRKHYTKNLLTFECHFDRGIDIDPLPIIQVEAGS 1108
QY 772 --KTDAASNKLIAMKTEKEMSFEPDDQSGSELNLEELIDDLQNSQLPOLFPDTRPGA 829
DB 1109 KKKTAASAASVSPSGSSNSQDSPAPPSAPMAIDG-----YPGYPGGS 1152
QY 830 P-----AGSVQKQALINDLMQTLAENSPLYVGAQKALRISOSTFNNRPPGQGLPLPNQ 885
DB 1153 PLYVASPPQPDYATAGOMQRPSPONNPOT-----PHFGAAVAAYACD 1194
QY 886 NLPLDITLQSPGTAGFPPIRNSPSYVIPQPG--MKNQGMIGNOGNLGNSGTGMSGNSA 944
DB 1195 NISVSNFEDPIPAAGGPGSGTGRPGQGGPRFAASGAAVAVAG-----GCPQ 1244
QY 945 SRPTMDSGEWAPQSSAVRYTCAATTSANMRPVGGMIRNPASIPMRSSQPGQROTQLOS 1004
DB 1245 PHPPRP--SPRT-----AAQQAAGQHQOONPHQNHGRLGPRPPRQOQOQCG--QQ 1292
QY 1005 QVANNIGSELEMMGGRQYSSQQAAPRQGTAPWES--ILPIDAQSFASQNNQPRPGSSPDDL 1063
DB 1293 PPSVVG-----GGPPAPQOHGPGQVPPSPQOHNHRAAGAY-----PPGSGSYPTP 1339

```



**THIS PAGE BLANK (USPTO)**

[illegible]



Query Match	100.0%; Score 7631; DB 19; Length 1464;
Best Local Similarity	100.0%; Pred. No. 0;
Matches 1464; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
1 MSGMGENTSDPSRAETRRKRECPDOLGSPKRNTERKRNQENKXYIEELAEIIFANFNDI	60
1 msngmgentdspraetrkrkecpddqgpeprkntekrnegemkyieelaelfanndi	60
1 msngmgentdspraetrkrkecpddqgpeprkntekrnegemkyieelaelfanndi	60
61 DNEFNKPPKCALIKETVQIOIRIKQEKRAANIDEVQKSDVSSRIGOGYIDKDALGPMML	120
61 dnefnkppkcaliketvqioirirkqekraaanidevqksdvssrigogyidkdalgpmm	120
61 dnefnkppkcaliketvqioirirkqekraaanidevqksdvssrigogyidkdalgpmm	120
121 EALDQFFEVFNLEGNVFEVSENVTOYLARNQELMNRKSYSTIHVGDHTEFPKLLPKSI	180
121 ealdqffevfnlegnvfevsenvtoylarnqelmrksystihvgdhtefpknllpkxi	180
121 ealdqffevfnlegnvfevsenvtoylarnqelmrksystihvgdhtefpknllpkxi	180
181 VNGGWSGSEPPRRNSTTENCRLVNRPLDPSSEEGHNDENAHOKYETMOCFAVSOPKSI	240
181 vnggwsgepprrnsttencrlvnrpldpsseeghndenahokyetmocfavsopkxi	240

Db	181	vngswsgpprrnshtrfncrmlykxlpdpdeeeahngeahkyetmqcfavsqpksike	240
Qy	241	EGEDLQGLCIVARVPMKERPVLPSSSEFTTRQDLQKITSLDTSYMKRAAMPQMEDLV	300
Db	241	egedlgsclicvarrvpmkerpvlpssefttrqdlqgktsldtsymaakpjewdlv	300
Qy	301	RRCLOKTHAQHEGGSVYAKRHHHEVLRQGLAFSQYTRRSLSDDTLVAAQYTKSKILRSQT	360
Db	301	rrcloqkthaqheggsvyakrhhhevlrqglatsqytrrsldstlvaqtkskilrsqt	360
Qy	361	TNEPOLVLSJHMLHREONVCMVMDLTGCMGPKLPISSNSPAHQALCSGPGQDMULS	420
Db	361	tnepqlvlsjhmhreqnvcmvmdltgctumgkplnpsnsbpaqalcsqpgqdmuls	420
Qy	421	SNINEPINCPEQMGMGPRFGSGGMNHSGQATPCGGSYALAKMNSPSQSSFGMNP	480
Db	421	sninfpincpkeqmgmmpfgrsggmnhysgmqatclpgsgsyalkmnspsqssfgmnp	480
Qy	481	QPTSMLSPRHMSBPVAGSFRIDPSPGSPAGSLHSPVGCSTGNSHSTYNSLALQAL	540
Db	481	qptsmisprhmsbpvagsfridppsgfspagslhspvgcstgnshtynslalqal	540
Qy	541	SEHGCVSLGSSLASPOLKMNLONSVMNMPPLSKMSGLSDKDFGLXGEFSBETTQGA	600
Db	541	seghvslgsslaspolkmnlonsvnmpplskmsglsdckdfglxgefsbettdga	600
Qy	601	ESSCHPGOKETNDPNLPVAVSERRADGOSRLHDSGCRKLLQOLLTTTSDQMEPPLASS	660
Db	601	esschpgoketndpnlpvavsserradggsrlhdsqgrkllqllttsdqmepplass	660
Qy	661	LSDTNKNOSTGSLPGSGTHGTSLEKHKHKLHRLLODSSSPVDLAKRLAEATGKDLSDSS	720
Db	661	lsdtknkstgslpgsgthgtslekhkhllhrllqgsspvdalakraleatgkdlsgss	720
Qy	721	STAAGSEVTTIKQEVSVKKEKNMLKXLLDKDXTKRIQIGPETPKLERLDSTDPASNTK	780
Db	721	stapgsveitkqevsvkkkenallxylldkddtkigipeltprkerldstkdpsantk	780
Qy	781	LIAKKTEKEEMSEFGPOPSELNDLBEELDLQNSQLQLEPPDRPGAPASVYKQAIT	840
Db	781	liakktekeemsefpgpdpseldneellddlgmsqlpqlfpdrtrpapasvdkqait	840
Qy	841	NDLMLQTLAENSPLYVGAQKOTALRISOSTFNNRPGQLRLLPNQNLPLDITLQSPGAG	900
Db	841	ndlmlqlaensplyvgaqkltalrllsqtfnmrrpgqlgrllpnqnlpldillqspgag	900
Qy	901	PPPPTRNSPSTVYLPQGMNGNGMIGNGNLGNSSTGIGMSASRPTMPSSEMAPOSA	960
Db	901	pfpplrnspsvylpqgmngngmignqngnlgnastcgmignasrptmpssewapsa	960
Qy	961	VRVCAATTSAMNRPVGGMIRMPAASIPRPSPSGOQOTLOSOMNITGSELEMMNGG	1020
Db	961	vrvcfaattsamnrpvvggmrlmpaasipmrpspqgqqlcrlqsgymnlhgslelmmngg	1020
Qy	1021	POYSOQOAPNOTAPWESILLPIDQASFSQONKOPRPGSSPDDLLCPHAAESSDEGALL	1080
Db	1021	pyysqdgapnotapwepesillpidqgsfasqnrpqrgysppddlcpbpaaesdegal	1080
Qy	1081	DOLYTLARNFDGLIEDRALGIRELYSOSQANDPEQFSSQSDNSIMLEOKAPVTFPOQVASQ	1140
Db	1081	dqlytlafrnfdgleedralgirellyslvsqsdavdeqfssqsdnsimleqkavrfpqvasq	1140
Qy	1141	AOMOGSVPMDONPHTMGORPSTVTLNRQPPGIRPRLGVLONONPOLRLQLOHRLQAO	1200
Db	1141	aomogsvpmddnphntmgorpsvtalnrqppgirprrlgvlononpolrqlolohrlqao	1200
Qy	1201	QNRQPLNKOJNSVSNVWLTLRPVCPVQAPLINAQMLAQORREILINQHLRQRMHQOQOQYQ	1260
Db	1201	qnrrqplnkojnsvsnvwltrpvcvpqaplinaqmlaqorrelinqhlrqrmhqoqqyq	1260
Qy	1261	RTLMMRQGLNMTFSSNVAFSGMDATMSNPRILPQANAQOFPPPPNYGISQODPDCFTGATY	1320

Db 1261 rllmmrggglmltpsmvapsmpatmsnprilpqaagqfipfpnyqisqppdftgattc 1320  
Oy 1321 POSPLASPRMAHTOSPMQOOSANPAYOAPSPDINGMAOGNMGNSMPQSPPHGGOAN 1380  
Db 1321 ppsplmsprmahltqpmqgspqanpayaqpdlingwaqgmngmsmtsqspbhqgqan 1380  
Oy 1381 TSMYSNNNNINVSMAATNTGMSNNQMTGQISMTSVTSVSTGLSSMGPEQVNDPALRG 1440  
Db 1381 tsmysnnnninvsmaatntgmsnnmqmtqgqismtsvtsvstglssmgpeqvndpalrg 1440  
Oy 1441 NLFPMQLPGAMIKQEGOTTAKYC 1464  
Db 1441 nlfpmqlpgamkqegdtlrykc 1464

RESULT 2  
AAM99482  
ID AAM99482 standard; Protein: 1463 AA.  
AC AAM99482;  
XX  
XX 08-JUN-1999 (first entry)  
DE Murine NCoA-2 protein.  
XX  
XX Mouse; P/CIP; p300/CBP/co-integrator-associated protein; gene expression;  
KW breast cancer; inflammatory disease; atherosclerosis; osteoporosis.  
XX  
OS Mus sp.  
XX  
FH Key Location/Qualifiers  
FT Misc-difference 989  
PD 17-DEC-1998. /Label= unknown

XX MO8656806-A1.  
XX 17-DEC-1998.  
XX 12-JUN-1998; 98WO-US12263.  
XX 12-JUN-1997; 97US-0049452.  
XX  
XX (REGC ) UNIV CALIFORNIA.  
PI Glass CK, Rose DW, Rosenfield MG, Torchia J;  
XX WPI: 1999-080883/07.  
XX  
XX New nucleic acids encoding P/CIP and NCoA-2 polypeptides - are used  
PT to identify agents that regulate gene expression, e.g. for treatment  
XX of cancer, inflammatory disease and osteoporosis  
XX  
XX Claim 22; Fig 2A; 100pp; English.  
XX  
XX This sequence represents the amino acid sequence of the mouse NCoA-2.  
CC The protein can regulate gene expression so are potentially useful  
CC therapeutically, e.g. against (breast) cancer, inflammatory disease  
CC (e.g. atherosclerosis) or osteoporosis.  
XX  
SQ Sequence 1463 AA:

Query Match 78.2%; Score 5969.5; DB 20; Length 1463;  
Best Local Similarity 80.5%; Pred. No. 0;  
Matches 1191; Conserved 95; Mismatches 163; Indels 33; Gaps 9;

Oy 121 EALDGEFFVNLNENWVESENVTOLRYNOBELANKSVSTLHVGDHTEPVKNLLPKSI 180  
Db 121 kalgffivnlengsvvfinvtqyllyngqklnmksvstllnygdclvtvknllpxsm 180  
Oy 181 VNGGSMSEPPRRNSHTNCRMLVRLPDSEEGHDNDQAQKYETMOCFAVSQPSIKE 240  
Db 181 vnggswsepprrnshtntcmrlvklpdsckkngndspkamykamcfavsqpsike 240  
Oy 241 EGEDLOSLT-----CYARVPKEREVLPDSSEFTTRDLOGKITSLDYST 287  
Db 241 egedlqslivmedqphgkltmslirk-lynpqpprdqmdltghyme-----sr 290  
Oy 288 MRAAM-KPGWEDLVRCLOKFEHQEGESVSARAKHHHEVLRQGLAFQIYRFSISDCTL 346  
Db 291 heagligresqkdatf-----stshn-kslylmprthnevlrqlatlsqylfslsdqcl 346  
Oy 347 VAAQTKSLIRSQITNEPOLVLSLMLHREQNVCMVNPDLTGOTGKPLNPISSNSPAHQ 406  
Db 347 vaagtksklirsqtneqqlvislhlhreqmncvmpdltgqangkplnplssspahq 406  
Oy 407 ALCGSPGQDMTLSSNINFPINGKEQMGMPNGRFGSGGMNHWGMOATTPQGSNTALX 466  
Db 407 alcspnpgqdmthlssnhfphngkqghbqtrfgsggmngvsgbqattpqgsnyalk 466  
Oy 467 MNSPQSSPGMNPQGTSMILSPRHRMSPVAGSPRIIPSPQSPAGSLSPVGVCSGTGNS 526  
Db 467 mnspsqsspgmnpqgavsvlsprqrmspvagsprlpspsqspagslspvgyvcsstgns 526  
Oy 527 HSYTNSLNALQALSEGHGVSIGSSIASPDLMGNLQMSPVMMNPNPLSKMGLSDSKCF 586  
Db 527 hsythsnalqalseghgvsigssiaspdlkngnlqmspvmmnhpprlskhsglsdskcf 586  
Oy 587 GLICEPSECTTGOANSSCHPGQKETTNDPMLPRAVSSERADQSRHLDKQTKLLQLLT 646  
Db 587 glycepsecttgoansschpgqkettndpmlpnavsseradqsrhldskqtkllqlt 646  
Oy 647 TKSDQMEPSPLASSLSDTNKDSGTSLPGSGNHGSLTEKKHILHRLQDSSSPDLAKL 706  
Db 647 tksdqmepsplasslsdtnkdsgtsglpgsgnhgsltkkhlhrlldqssspdlakl 706  
Oy 707 TAATGKDLQSESSSTAPGSEVTIKQEPVSPKKENALLRLVLDKDDTKDGLPEITPKL 766  
Db 707 tahatghlsgessstapgstevtkqepvspkkemallrlyldkddtkdglpeitpkl 766  
Oy 767 ERLDSKTPASNTKLIAMKTEKESFEPGQPSGELDNLEIILDDLQNSQLPOLFEDTR 826  
Db 767 erldsktpasntklianktivkeevsfepsgdpsklnlkkllldlqhsqqlfepdtr 826  
Oy 827 PGAPAGSVDKQATINDLQTLAENSPPVPVGAOKTALRISQSTFNPNRPPGQLRPQN 886  
Db 827 pgapagsvdkqatindlqtladssppvpaqgkaalchsgsstnmppgqlrplpqn 886  
Oy 887 LPLDITLOSPTGAGFPFPIRNSSPYSVIPQGMGNQGMIGNOGNLGNSGTGMIGNSASR 946  
Db 887 lplditlpspgagpfpirnspsysvipqgghngmlsgnlgphstghshgsestr 946  
Oy 947 PTMPSGEWAPQSSAVRVTCATTSAMNRPVOGGMIR--NPAASIPMRSSPGORQTLQS 1004  
Db 947 ptmpsgewapqssavrvtcattsamnrpvoggmir--npaasipmrsspgorqlqts 1004  
Oy 1005 QVMNIGPSELEMMNGGPGYSSOQOAPPNOTAPWPSIILPIDASRASONRQFFGSSPDLL 1064  
Db 1006 qvmnigpselemmnggpgyyssoqoappnotapwpsilpidasrtasqnrqffgsspdll 1064  
Oy 1065 CPHPAESPDEGALLDQLYALRNFDLEIIDLALGPELVVSQOAVDPQFSSQSDNI 1124  
Db 1066 cphpaespsdegalldqlyalrnfldleidlalgpelvsqavdaefssqsdni 1124  
Oy 1125 MLEQKAIVFPQOYVSAQMAAGSYSPMODPNFHTMGORPSTATLRMOPRGLRPTGLVON 1184  
Db 1126 hleqkavfpqoyvasaqmagyspmodpnfhtmgorpsatlrmpogrlrptglvqn 1184

QY 1185 QPNOLRLQHLQRLQAOQNPRLMNQISNVSNVLTLPVPTQAPINAQMLAQORREILN 1244  
 CC formation of SRC-3 transcription complexes. Determining the amount of  
 CC SRC-3 in a sample is also useful in a method to detect overexpression of  
 CC SRC-3 and therefore determining the neoplastic states of cells in humans  
 or animals.  
 Db 1186 qpqqlrlqqlrlqagqnrqlhmqssvsnvnlrlrpyvtqaplnagqlagqrqlh 1245  
 QY 1245 QHRLQRMHQOQVOQRTLMRGQGLMTPSMWAPSGMXPATMSNPRTRPQANAOFPPPN 1304  
 Db 1246 qhlrtqy--hggqvqrqlhrlhrgqllhvrpslvapaglpashhprlpqahqqfipflph 1303  
 QY 1305 YGISQOPDPFGTATTPOSPLMSRMAHTQSPMMQSOANPAYOAPSDINGMAQNGMGN 1364  
 Db 1304 yglsgqdpdpgfqtatcpqslhspthahcgsphhqsqanparyqpsdhdngvqsgshgn 1363  
 QY 1365 SMFSQGSPPHFGQOANTSMYSNNMNVSNATNTGKSSNMNQTGOISMTSVTSVSTGL 1424  
 Db 1364 shfsqgspphfgqantshysnnhlsvshatntglsahhqtgshstsvtsvpsgl 1423  
 QY 1425 SSMGPEQVNDPALRGGLFNPQLPGMDMIKQEDDTPPKYC 1464  
 Db 1424 pshgpeqvndpalrggllfpnqllghnhlkxqegdastrkyc 1463  
 RESULT 3  
 AAY21975  
 ID AAY21975 standard; protein; 1522 AA.  
 AC AAY21975;  
 DT 06-SEP-1999 (first entry)  
 XX  
 DE Human steroid receptor coactivator-3 (SRC-3).  
 KW Steroid receptor coactivator-3; SRC-3; transcription enhancement; human;  
 hormone response element; estrogen receptor alpha; neoplastic state.  
 XX  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT MISC-difference 1424 /note= "encoded by TGA"  
 FT MISC-difference 1432 /note= "encoded by TAA"  
 FT MISC-difference 1474 /note= "encoded by TGA"  
 FT MISC-difference 1510 /note= "encoded by TAG"  
 FT MISC-difference 1521 /note= "encoded by TGA"  
 FT MISC-difference 1521 /note= "encoded by TGA"  
 PN WO9932621-A2.  
 XX  
 PD 01-JUL-1999.  
 XX  
 PF 01-DEC-1998; 98WO-US25478.  
 XX  
 PR 22-DEC-1997; 97US-0066511.  
 PA (AMHP ) AMERICAN HOME PROD CORP.  
 PI Fraill DE, Lyttle CR, Suen C;  
 XX WPI: 1999-405171/34.  
 DR N-PSDB; AAX80992.  
 XX  
 PT Nucleic acid encoding a steroid receptor co-activator-3, useful for  
 determining the neoplastic states of cells in humans or animals  
 XX  
 PS Claim 2, page 61-68; 75pp; English.  
 CC This represents a human steroid receptor coactivator-3 (SRC-3). Host  
 CC cells transformed with vectors comprising the SRC-3 gene can be used for  
 CC the recombinant production of the SRC-3 protein. SRC-3 can be used to  
 CC enhance the transcription of genes, which are regulated by hormone  
 CC response elements, e.g. estrogen receptor alpha. SRC-3 can be used in

CC methods to screen for compounds that bind to SRC-3 or which modulate  
 CC formation of SRC-3 transcription complexes. Determining the amount of  
 CC SRC-3 in a sample is also useful in a method to detect overexpression of  
 CC SRC-3 and therefore determining the neoplastic states of cells in humans  
 or animals.  
 SQ Sequence 1522 AA  
 Query Match 35.9%; Score 2739.5; DB 20; Length 1522;  
 Best Local Similarity 42.2%; P-Val 4.5e-170;  
 Matches 648; Conservative 243; Mismatches 453; Indels 191; Gaps 44;  
 QY 1 MSGGENTSDPSRAEPLKQECEDPOLGSPKRTTERKRNROEKYIEELAELEAFNNDI 60  
 Db 1 msglgenl-dpladsarkrlpdcptpgqgltcsgekrtrregeekylelaelsanlsdt 59  
 QY 61 DNFNFKPKCALIKETVQKIQRIKEQEKAAANIDEVQSDVSTGQVIDKALGPMML 120  
 Db 60 dnfnvfkpkcaliketvqrqikegk-rlsnddvqkadavstgqvvdkslplll 118  
 QY 121 EALDGFEEVNVLEGNVVFESENTOYLRYNOELMKSVYSTLHVGDHMEFVKNLPRK 180  
 Db 119 qaldgflfvnvgdnlvfyseavnlqykykqedi vltvsyvlhheedrlfxn-1pkst 177  
 QY 181 VNGGSWGEPPRRNSHTFNCRLMLVKPLPDEEEDHNOEAHQYETMOCFAVSQPSIKE 240  
 Db 178 vngvswtneqrqkshntfncrlmlktrphdlledinspennqrqyelmqcfalsqpramme 237  
 QY 241 EGEDLOSCLICVARRYPMKERYLPSESEFTTQODLOGKTTSDTSTMAAKPGMEDLV 300  
 Db 238 egedlscmlcvarritlterftrpnefclthldslgkvvndtlnslrsmptgfedll 297  
 QY 301 RRCIOKFAHQEGESVYAKRHHEVLRQGLARSOIYRFSIDGTVAQTKSLRSGT 360  
 Db 298 rrciqrflfslndgsws-qkrlhyqeaylmghecpvyrflsldgltvtaqsklfrnpy 356  
 QY 361 TNEPOLVISLMLRLBONCVNPNDLTGOTMGRPLNDPISNSPDAHQALSGNPDQMTLS 420  
 Db 357 tndrhgfvstchflqregnqyrpnpvgglrprmgacnssvsg----msmprngllqmp 412  
 QY 421 SNTNFPINGKEQMGKPMKRGFGSGGNNHVS---GMOA-TTPGGSNYALKMNSPSSSFC 476  
 Db 413 ssrnyladpsttgmsgagayssnlasltpbgqpspsyyqnnqylmmspspbgspg 472  
 QY 477 MNPQPTSMSPRRHMSPCVAGSPRIIPQSPFAGSLHSPGVGCSSTGSHSTNSLNA 536  
 Db 473 lnpnqnlmsprnr-----gspklashqfispvagvshpmaassnltgn hstssssla 525  
 QY 537 LQALSEGHVSLGSLASPDLMKGNLQNSFPVNNPPRLSKMGSLDKDCFGLYGE--PSB 594  
 Db 526 lqalsegyvtslslstlsspprk---ldnsp-mnmltqpskvnsqdeksprgfydcqdnpe 581  
 QY 595 GTTGQAESSCHPEQKQETNDPNLPPAVSSERADGOSRLHDSKQTYLDLTTKSP---- 650  
 Db 582 smtcqnsrldhsl-d-keskes-----svgeaenqgrplekshkklqlldtsssdrgn 634  
 QY 651 -QMEPSPLASSTL-----SDTNKDSYGSPLGSGSTGTSLKXKKNLHNLDDSS 699  
 Db 635 ssitnspldssckessvvtspgvsstcgayssstnmhgsllqekhlthlnldngns 694  
 QY 700 PYDLAKRLAETATGKDLQESSSTAPGSEVYTIKQEPVSPKKE-NALRYLLDKDPTKDG 758  
 Db 695 paevaklteatgkd--ssltscgdnvnykqeqslspkknennallrylldtdpdsdal 751  
 QY 759 LPETITPKLERLDSKTPDASATKLIAMKTEKE-EMSPGPCQDQSELDNLEETIDLDQNSQ 817  
 Db 752 skeltpqevdnlkmsqcsstlpsssqekprikleteesegydldnldaldgtlssd 811  
 QY 818 LPQLPDTRGAPAGVVDKQAIINDLMQLAEKSPVTPVGAQKTARISQSTFNNRPPQ 877  
 Db 812 lynnsssn-gshlgt--kqvvtg-----gtlslgkssqsv-qsltrp-- 850



```

Db      357  tndrhgfvstchfgrdqnyrpnbpvgqgllrpfmagcnssvvg-----mmsmpqglqnp 412
Qy      421  SNINPFINPKKMGKMPMGKFGSGGMNHVS---GMOA--TTPGNSVYALKMNPSSQSPG 476
Db      413  sstrayladdpstcgqmsgarvysnsiaaltgpgmgspssygmnyglmmsapbpgspg 472
Qy      477  MNGOPTSMLSPRHMSPGVAGSPRIIPSOFPDAGSLHSPVGCSTGNSHSTYNSSLNA 536
Db      473  lapnqgnlmsprnr-----gsplksnhqfsvpavghspmaasgnlgn- hsfssslsa 525
Qy      537  LQALSCHGVSLGSSLASPDLMKGNLQNSPVNMPPLPMSGLSKDCKDFGLXGE--PSE 594
Db      526  lqaisegvtslslstlsspgpk---ldnsp-mmltcqskvsnqdsksplgfycdqnpve 581
Qy      595  GTTGOAESCHPEGEKQETJNDPNLPRAVSSERADQSRHDSKQOTLLOLTLTKSDP---- 650
Db      582  smcgsnsrdhlsd-keskes-----svegaenqrgplskghkhlldlscsddrgh 634
Qy      651  -QMEPSPLASL-----SDTKNDSTGLPGSGSTHGTSLKHKHKLHRLLODSSS 699
Db      635  ssitnpldssckesvsvtlpsgvssstsgvssstsmhgsllgkhrllhkllqngns 694
Qy      700  PVDLAKLTAEATKDLSDQSSSTAPGSEVTIKOEPVSPKKKE- NALLRITLKDQDKDG 758
Db      695  paevakltaeatgk-d--tsstscgdnvkvqegqlspkkennallrylldrddpsdal 751
Qy      759  LPEITKLELSDKTDTPASNTKLIAMKTEKE-EMSEFPGDQPESELDNLEIILDDQNSO 817
Db      752  skelqgvqevdnhkmsqclsstlpsssqekdkiktetseesgddndldatgdtssd 811
Qy      818  LPQLPDTBRGAPAGSVDKOAILINDLMQLTAEKNSPYTPVGAQKALTALRISOSTFNNRPGQ 877
Db      812  fymslssn-ghlgtf--kqvvtg-----gtneislkssqsv--qslrtp-- 850
Qy      878  LGLLNLNQLPLD--ITIOSPTGAGPFPPIRNSPYSVIP-QCGMMGNQCMIGNQNLGN 934
Db      851  -----pytravslspvsvsgspvknlsatfmlpkqgmqlgnptmmdsgnys 900
Qy      935  SSTGMIGNSASRPTMPSGEWA-POSSAVRYTCAATTS-----AMNRPVQGMIRN 983
Db      901  smgpgprntvtctgssgdglpnskaqgmegmmsmrgpssgdytslpralgsi-- 958
Qy      984  PAASIDPMRPSSQPGQRQTL--QSQVNNIGPSELEMMNGGFPQYQQAAPPNQTAPWPESTL 1041
Db      959  --plbfrlrsnslpgarvblqgqgmqlmrgpgelmgmganpyg--qaaasqlgswedma 1015
Qy      1042  PIDQAFSAQNRQPFSSPDLLCPHPAESPSDEGALLDQLALARNPD--GLEIEDRA 1099
Db      1016  smeqvshgicqnrpllnrslddlvgppsnleggsderallldqhlntlsntdaqlseidra 1075
Qy      1100  LGIPELVSOQAADPEQ--FSSQDSNIMLEOKAPVPPQQAQAQAQAQSSYSPMODPNFH 1157
Db      1076  lglpelvngqgallepqdaftgqgeaavmmdqkaglygqfypaagppmgq-----gfh 1127
Qy      1158  TMGQRPSTVATLRMO-----PREGLRPTGLVO---NOPNQLRLQLHRLAQO---QN 1202
Db      1128  lqgqsfmsfsmmnmqmgqgnfplqgmhpranlmrptcnlpkqlrmglqglqgqgflngs 1187
Qy      1203  RQPLMNOISN-VSNVNLTLRPGV-PTQAPRTNAQMLAORQELLNOHLR----- 1248
Db      1188  rgalclkmempftaggaavmrpmmqpgqgflnaqmvagrstellsbhftrqrvammngqgq 1247
Qy      1249  -----QROMHQOQOVQOQRTLLMNRQGLMTRPSNVAAPSCMPATMSNRPANQAOQFPE 1301
Db      1248  qggqgqgqgqgqgqgqgqgqgcqafspppnvtaaspmmdglagpctmpaqpqpqfpy 1307
Qy      1302  PPNYGISQOPDPGFGATTTPQSPFLMSPRMAHTQSPMMAQSQANPAYQAQSDINGMAQGN 1361
Db      1308  qpnygmngqgqpdqafgyrvtsspnmmasrmngspqpmmqhpbqaslygs--semkyvpsgnl 1366
Qy      1362  GNSMSQGSPPHFGQOANTSMTSNNMNTINVSMAITVTGKSSNMQMTGOISMTSVTSVST 1421

```

```

Db      1367  arnsfsfq---qfabqgnpavys-----mvhnmngssghngmnmmpmpn 1408
Qy      1422  SGLSSMGPEQ 1431
Db      1409  sgm-pmgpq 1417

RESULT 5
AAB59278
ID AAB59278 standard; protein; 1412 AA.
XX
AC AAB59278;
XX
XX 27-MAR-2001 (first entry)
XX
DE Human SRC-3 protein.
XX
KW Steroid receptor coactivator-3; SRC-3; antisense; infection;
KW Inflammation; tumour; cancer.
XX
XX Homo sapiens.
XX
XX US6156571-A.
XX
XX 05-DEC-2000.
XX
XX 15-NOV-1999; 99US-0440612.
XX
XX 15-NOV-1999; 99US-0440612.
XX
XX 15-NOV-1999; 99US-0440612.
XX
XX (ISIS-) ISIS PHARM INC.
XX
XX Bennett CF, Cowselet LM;
XX
XX WPI; 2001-079549/09.
XX
XX Novel antisense compound useful to prevent or delay infection,
XX inflammation or tumor formation, specifically hybridizes with and
XX inhibits the expression of human steroid receptor coactivator-3
XX
XX Example 13; Column 43-54; 36pp; English.
XX
XX
XX The present invention relates to an antisense oligonucleotide,
XX targeted to a nucleic acid molecule encoding human steroid receptor
XX coactivator-3 (SRC-3). The invention is useful for inhibiting the
XX expression of SRC-3 in human cells or tissues in vitro. It is
XX useful for diagnostics, therapeutics, prophylaxis and as
XX research reagents and kits. It is useful prophylactically, to
XX prevent or delay infection, inflammation or tumor formation.
XX
XX Sequence 1412 AA;

Query Match 35.5%; Score 2711; DB 22; Length 1412;
Best Local Similarity 42.3%; Pred. No. 2,9e+168;
Matches 642; Conservative 242; Mismatches 439; Indels 194; Gaps 46;

Qy      1  MSGMGENTDPSRAETRRKKECPDOLGPPKRTKERNRNEOEKRYTEELAEILFANFNDI 60
Db      1  msllgenl-dplasdskrkklpodtpgqgltscegekrirreegskyleelaelsanlsdt 59
Qy      61  DNEFRPKCALIKETVKOIRQIKQEKAANAANIDEVQKSDVSTGQVYDKDALGPMML 120
Db      60  dnfnvxpdkcaliketvtrqglkegk-tisndddvqkavssstgvyldkslpglll 118
Qy      121  EALDGEFFVYNLEGNVVEYSENVOTYLRKNQOEFLMKSYSYLHGDHREPKNLPKSI 180
Db      119  qaldfllfvvndgnalvfsvenvtgylykqgedlvntsvynllheedrklfkn-lpkst 177
Qy      181  VNGSGSGSPRRNSHTFRCRMVVRPLPSEEGHNOGAHQKYMOCFAVSQPSISE 240
Db      178  vngsvwtneitqgkshfncrmmlktrphlledlnaspeimrgiryetumcfalsqpramme 237

```







PT steroid receptor coactivator-1, particularly useful for preventing,  
 PT delaying or treating infection, inflammation or tumor formation -  
 XX  
 PS Claim 1: Column 45-54; 36pp; English.

CC The present invention relates to an antisense compound of up to 30  
 CC nucleobases in length, which specifically hybridizes with and inhibits  
 CC the expression of human steroid receptor coactivator-1 (SRC-1) (also  
 CC known as F-SRC-1 and NcoA-1) gene. The antisense compounds are useful  
 CC for diagnostics, therapeutics, prophylaxis, or as research reagents or  
 CC kits. The antisense oligonucleotides are useful for treating a disease,  
 CC particularly a human, suspected of having or being prone to a disease  
 CC or condition associated with the expression of SRC-1. In particular,  
 CC the antisense oligonucleotides are useful for preventing, delaying or  
 CC treating infection, inflammation or tumor formation. The present  
 CC sequence is human SRC-1.

XX Sequence 1441 AA;

Query Match 28.3%; Score 2159; DB 22; Length 1441;  
 Best Local Similarity 36.0%; Pred. No. 4.2e-132;  
 Matches 579; Conservative 235; Mismatches 426; Indels 368; Gaps 61;

QY 1 MSGMGENTSDPSRAETRRKRECP-DOLGPPKRNTEKRNREKNYIEELAFANFMD 59  
 DB 1 msgjgdsdpsraetrrkrkrcpdlolgppekrrntekrrnreknyleelafanfm 56  
 QY 60 IDNRFKDKCALIKETVYKQIRKOEKAAANIDENOKSVSTGGCVIDKDALGPM 119  
 DB 57 idnsrvkdkckalktktdqqlkrmekexstdddvqsdsssgvlekesjgpll 116  
 QY 120 LEALDGFVVVNLGNNVFEFNVTOYLRYNOEELMANKSVSILHVGHTEPVKMLTK 179  
 DB 117 lealdgffvvnlgcnvfeenvtsylygnqeelmncsvsilhvghaetvknllpks 176  
 QY 180 IVNGSGMSGEPRRNSHTFNCRLMLKPLDSEEGHNOEAHQKYETMOCFASOPKST 239  
 DB 177 lvngsvpgeatrrnshfncrmllhp-pd-epgteneaqeyvmqcfvsgpklsiq 233  
 QY 240 EEGDLOGLCVARVPMKRPVLPSESEFTTRDLOGKITSILDTMRAAMKRGWEDL 299  
 DB 234 eegedfglclclartlrp-trpaltgyesimtkdqdtcgtklisidtslraagrgwedl 291  
 QY 300 VRRCIQKHAOHEGSESVYAKRHHNEVLROGLAFQIRFSLSDGLVAAGTKSLRSQ 359  
 DB 292 vrrcklyaff-qpgyregpsyaqlfgevmtrtlaaspsyrflldngctmsahckcklcy 350  
 QY 360 TTNEPQVLYSLHMLHREONVCMNP-DLTGOTMGRP-LNPISNSPAHQALCSGN- 412  
 DB 351 spdmqpfimgjhlidrehs--glspqddtngmslprvnpvnp--slspahgyarstsl 406  
 QY 413 -PGQDMTLSSINFPINPKEO-----MGMPIRGSGSGGMNHVSGM-----QA 455  
 DB 407 ppsnsmvstcin-----rqgsdilhssshnsnsgsfqcpqsglvanvalnkqga 460  
 QY 456 TTPGGSNATKMNPSQSGPMNPQPTSMLSPRHRMSPGVAGSPRPSOPSP-AGSLH 514  
 DB 461 ss-qskspslhlnmpmegtgislaq---fmsprtrqvtslglatrlprmnnsfprpistls 516  
 QY 515 SPVGVCS--TGNSHSTYNSLSNALQALSEGHSVLSGSSLASPDLMGNLSPVNMNP 572  
 DB 517 spvgmstasacnnmrnsynlrvtslqgmegnpnsvfsgsaasplrmqsgsperlnd 576  
 QY 573 PLSMGSLSDSKDGLGVESEGTTGQAESCHPEQKET-----NDPMLPRAVSSERADG 628  
 DB 577 p-akaeskdmeleasclnemi-----qsdnssdckpklidsgllhndel1-----sdg 622  
 QY 629 QSRLHDSKGFKLQLLTTKSD-OMEPSPILASSLD-----TKKDTGSLPGSGSTHG 680  
 DB 623 dskf--qstskhlyqlltttaeqglrhaditckdvscgtstsaanssgsgcspsh 680  
 QY 681 TSLKEKHILHRLIQDSSSPVDLAKLTAEATGKDLSOESSSTAPG--SEVTIKOEPVS 737

DB 681 saltarhklhlrlige-gpsdlttltsvpekdckd-sastsavtqvgvnsksklelds 738  
 QY 738 KRKE--NALRYLDDKD-----TKDIGLPETPKLERLDKTDPASWTKLIAMKTEKE 789  
 DB 739 krkesdhllyllyldkdekdlrsipnlslddkvkvexke-qmdepcntprcmkcppe 797  
 QY 790 EMSFERGDDPGSELDLLEITLDDLN-SQLPOLFPDTRRGAGASVDKQAIINDLMQLTA 848  
 DB 798 eikleaqsqfidalqfdqdlpblekaaqipglcetdrnd---gavtsvlikxel----- 849  
 QY 849 EMSVTPVGAQRTALRISGSTFNPNRPGLRLPQNQLPLDITLQSPGAGFPPIRMS 908  
 DB 850 -----lpslqsat----- 858  
 QY 909 SPYSYIPIPGMNGMGNOGMIGNOGLNCSGTGMGNASRPTMPSGEWAPQSAVRVCAAT 968  
 DB 859 -----arp----- 861  
 QY 969 TSAMNRPVGGMIKRNPAISIPRPSQPOGROTLQSOVNVNIGPSELEMMGPOYSQQA 1028  
 DB 862 tsrlmr-----lp-----eleleaingfygpyt 885  
 QY 1029 PENQTAAMP-ESILPIDQAFASQNRPGSSPDLLCPHRAESPDSGALLDOL-VL 1085  
 DB 886 --gdqiprwnetvtaing--sksedqclsqldellcpptlvegrndekallleqlvsfl 940  
 QY 1086 ALRNFDELEEDRALGIPELVSQSOAVD--PBOFSSQDSN--IMLEOKAPVPPQVQASQA 1141  
 DB 941 sgkdelelaeldralgldklv-gggglldvlserfppqgqtrpplimeerpnlysqyspf 999  
 QY 1142 QMAQGSYSPMDQPNFHTMQRPSYATLRMQRPP-----GLRPTGLVON--QPNOL 1189  
 DB 1000 pta-----lpsrfqgmvrqkpslqtmprvtrp-rgafsgpmgmprqlintrpaarpql 1054  
 QY 1190 RLQLOHRLQAO-----QNRQPLAISVNSVNLTLRLPQ-----VPTQAPRTNAQMLAQR 1240  
 DB 1055 rqlqgrlqgqqqlhqnqgalnqfaatavrglmrgsmgqqlcpqpplnsgmlaqrg 1114  
 QY 1241 ELLNQHRLQRHMQOQOQOQRTLMRGO--GLNMTPSNVAIPSGMPTMSPRIPOANAOQ 1298  
 DB 1115 elysqghvqrqllyq-----qramlmrqsfignlpps-----sglpvqitgnprlpqgappq 1166  
 QY 1299 PEPFNPVGSQOPDQPGFG-----ATTPQSPV-----MSPRNA 1331  
 DB 1167 fpyppnygtupgtppaslsfsglaanpeaslaanrnmvsrgmtgnlgqfgtgnpqnq 1226  
 QY 1332 HT-----OSPMQOQOANPAYQAPSDINGM 1356  
 DB 1227 qnvfygpgagmvpqgeanlaspsssmvmpirppqgsllqgfprrasgyqsp-dmkaw 1285  
 QY 1357 AOGNNGSMFSQ--QSPRHGQOANTSMYSNMNMNINVMATNTGSGMSSMNQTOISMT 1414  
 DB 1286 qggalgnnnvtfqavqnpq---tpaqpyvy--dnmsiltvmaqgnltvnygmnmmaqmgn- 1340  
 QY 1415 SVTSGTSGLSMGPEOVNDPALRGNLFPNOLPGMDMKOGGDTTRK 1462  
 DB 1341 --stlqpmgmvtlvceqindpalrntgilycnqjstcdllktaadgtq 1366

RESULT 8  
 AAM99481  
 ID AAM99481 standard; Protein: 1362 AA.  
 AC AAM99481;  
 XX  
 XX 08-JUN-1999 (first entry)  
 DE Murine p/CIP protein.  
 XX  
 XX Mouse; p/CIP, p300/CBP/co-integrator-associated protein; gene expression;  
 KW breast cancer; inflammatory disease; atherosclerosis; osteoporosis; ds.

OS	Mus sp.	Location/Qualifiers
XX	Key	Location/Qualifiers
FH	Misc-difference 540	/note= "defined in specification as ?"
FT	Misc-difference 1080	/note= "defined in specification as O"
FT	Misc-difference 44..45	/note= "there appears to be 20 amino acids missing between these positions in the sequence given in the specification"
FT	Misc-difference 1044..1045	/note= "there appears to be 20 amino acids missing between these positions in the sequence given in the specification"
FT	Misc-difference 1044..1045	/note= "there appears to be 20 amino acids missing between these positions in the sequence given in the specification"
XX	17-DEC-1998.	
XX	12-JUN-1998;	98WO-US12263.
XX	12-JUN-1997;	97US-0049452.
XX	(REGC ) UNITV CALIFORNIA.	
XX	Glass CK, Rose DM, Rosenfield MG, Torchia J;	
XX	WPI: 1999-080883/07.	
XX	N-PSDB; AAX26000.	
XX	New nucleic acids encoding p/CIP and NcoA-2 polypeptides - are used to identify agents that regulate gene expression, e.g. for treatment of cancer, inflammatory disease and osteoporosis	
XX	Claim 16; Fig 1; 100pp; English.	
XX	This sequence represents the amino acid sequence of the mouse p/CIP (p300/CBP/co-integrator-associated protein) polypeptide. The protein can regulate gene expression so are potentially useful therapeutically, e.g. against (breast) cancer, inflammatory disease (e.g. atherosclerosis) or osteoporosis.	
XX	Sequence 1362 AA;	
SO	Query Match	27.9%; Score 2127; DB 20; Length 1362;
	Best Local Similarity	39.2%; Pred. No. 4.7e-130;
	Matches 543; Conservative 214; Mismatches 424;	Indels 204; Gaps 44
QY	1 MSGMGENTSDPSKAETRRKRECPDQGPSPKRTNRRNREQENKYLELAELIFANFNDI 60	
DB	1 MSGMGENTSDPSKAETRRKRECPDQGPSPKRTNRRNREQENKYLELAELIFANFNDI 60	
QY	61 DNEFKPKCALIKETVVKOIRQIKEGEKAAANIDEVQSDVSSTGGVTDKALGPML 120	
DB	61 DNEFKPKCALIKETVVKOIRQIKEGEKAAANIDEVQSDVSSTGGVTDKALGPML 120	
QY	121 EALDGEFFVNVLEGNVVFSENTOYLRYNOELMLMKSVYSLIHHVDHTEFVNLPKSI 180	
DB	121 EALDGEFFVNVLEGNVVFSENTOYLRYNOELMLMKSVYSLIHHVDHTEFVNLPKSI 180	
QY	100 qalldgllfvvndglnvfyseavtqylqyqgedlvtvsvstlhprrkdfllntyqnpql 159	
DB	100 qalldgllfvvndglnvfyseavtqylqyqgedlvtvsvstlhprrkdfllntyqnpql 159	
QY	181 VNGSGWSGEPFRNSHTFNCRMVLVFKLPDSEEGHNOEAHQKYEPMQCPAASOPKSIKE 240	
DB	181 VNGSGWSGEPFRNSHTFNCRMVLVFKLPDSEEGHNOEAHQKYEPMQCPAASOPKSIKE 240	
QY	241 EGEDLOSLICVARRPMKRPVLPSPSEFTTRDQLOGLTSDTSMRAAMKPGWEDLV 300	
DB	241 EGEDLOSLICVARRPMKRPVLPSPSEFTTRDQLOGLTSDTSMRAAMKPGWEDLV 300	
QY	301 RRCIOKFAHQHEGESVSYAKRHHHEVLRQGLAFSQTLYRSLSDGTLVAQOTKSKLRSQT 360	
DB	301 RRCIOKFAHQHEGESVSYAKRHHHEVLRQGLAFSQTLYRSLSDGTLVAQOTKSKLRSQT 360	
QY	276 rrcioqffsindgqsws-qtrhvgeayvbnhaetpyrfisladgtlvsactsklffnpy 334	
DB	276 rrcioqffsindgqsws-qtrhvgeayvbnhaetpyrfisladgtlvsactsklffnpy 334	

QY	361	TNEPOLVLSLMLHREONVCVMMNDLTGOTMGKPELPISSNS----	PAHQALCSGNPCOD	416
Db	335	ndrnfgrfstftbfllrqn-----	gyrpnrlpqdkxgtrppaagcgvsmnpqn	381
QY	417	MTLSSINIFEPINGKREDMGMPGFGGSGGMNHVS-GMOATPR--	OGSNVYALKKNNPSQ	472
Db	382	vqmgmsttyppdpnsntcgmnggaaygassvasltpqgslqpsstysqnsyylsmsspph		441
QY	473	SSPGMNPQOPYSMLSPKRRMSPGVAGSPRIPPEQFEPAGSLSHSPVGVCSSTSHSYTNS	532	
Db	442	gsppqjgppnqjlmisprnr-----	gsphnashqfspaagahsnmpgnsntg-shsfass	494
QY	533	SLNALQALSEGHVSLGSSSLASPDLMKGNLQNSPVMMNPPLSKMGSLDKCFCFLYGE-	591	
Db	495	slsaqlasegvysllstlsspgpk---ldnsp-nmlnsqpskvsqdxkxsp19lyceq	550	
QY	592	-PSEBTTCQAASSCHPEGEOKETNDPNLPPAVSSERRDGGSRLLHDSNGQIKLLOLTITSD	650	
Db	551	npvesvscqsnur-dpvkxkkeskes-----	sgvesetprplessyghkklqlltccssd	603
QY	651	QMEPPEPLASSDPNR-KDS-----	TGSLPGSGSHGSLKEKHKLHRLQ	695
Db	604	drgnshltnspldpcckdxsvsvtspesgyvsscsqvtvstsnvhsllqekhrllhk1lq	663	
QY	696	DSSSPVDLAKTLAETATGKDLISOESSSTAPSGSEYTIKOEVPSPKKE-NALLRYLLDKDDT	754	
Db	664	ngnspaeavaktlaetgkd-----	tstascsgegttrceqelqspkkennalllyllldrddp	719
QY	755	KDIGLEPTEPKLEKDSKTDPASTKLIAMKTEKE-EMSEFEDDQGSLELMLLETLDD	813	
Db	720	sdvrlakelqppadgsdsklsgcscstnpssgqekdpikletndevsgldldhldlqgl	779	
QY	814	ONSQLPOLFEPTRGAPAPVSKOAILINDLMQTLAENSPTVPYGAQKTALRLISQSTFNP	873	
Db	780	ts-----dfymnptrngphga-kqgm-----	agps19lir-----sp	812
QY	874	REGQLRLLPNQNLPLDITLOSPTGAGPPIRINSFYSYIP-QPCMGMOGNIQNL	932	
Db	813	qvsvgrvppryravsld-----	spvsgvsppvkhnvasfpglpqpllagngprmmdsqeny	868
QY	933	GNSSTGMITGNSASRPTMBSGMA-POSSAVRV-----	TCAATTSSAMRPVQGM	981
Db	869	g-amngprnrpvpnrptspsgdwglansrasmeplasspigrtgadysatlprpamgsav	927	
QY	982	RNPASIMRPSOOP-----	OROTLOSQVANNIPSELEMMNGP	1021
Db	928	-----p1r1rnr1lpgarps1qgqgqgqgqgqgqgqgm1qmt1gel1pmngvyn	983	
QY	1022	OYSQOAPBPNOTAPWPESTILPIDOAFASQNRQFSGSPDDLCLPHRAESPSDEGALD	1081	
Db	984	pys-pavsgnpgswpegmlsmegpprgsqnrplrlrnsldl-----	1024	
QY	1082	QLYLALRNFDELELDRLALGPELVSSQAVDEQ--FSSQDSNIMLEOKAPVFPQVAS	1133	
Db	1025	hfl1sntatq1de1dralq1sp1elvnvgq1leskqdv1fgqgeaa1vymndk1aa1lyg1t1pa	1084	
QY	1140	QAMQMGSYSPW-ODPNFHTM-----	GGRPSVATLBMOPRGL-RPGVLVQONQONPRLQD	1199
Db	1085	gpprlqggf1n1qgqgsp1nsmgm1sqgqsf1qgm1pr1ag1v1r1r-----	tn1rpk1q1rnm1j	1144
QY	1194	QHRLOAQ-----	ONRQPLMNQJNSVNSVNNLTLPREVPTQAEINAMQALQORRETLNHLRO	1244
Db	1142	qgtr1qgqgflng1srg1a1emk1enp1ag-t1vm1r1p1mp-q1af1n1q1ma1gq1k1rel1msh1h1q	1199	
QY	1250	ROMHQOQVOQRTLMNR--	GQGLNMTPSNVAPSGKPA1TNSNRPIQAOAHQOPPEPNTG	1300
Db	1200	gfm-----	ammusyp1pyq1a1f1spp1nr1a1spsm1g1v1a1gsamp1q1ppq1q1fpy1any	1250
QY	1307	ISQOP 1311		
Db	1251	tgppp 1255		



KW	Steroid receptor coactivator-1; SRC-1; molecular switch;
KM	gene therapy; transgenic animal.
XX	
OS	Homo sapiens.
PN	MO971037-A1.
XX	
PD	20-MAR-1997.
XX	
PF	20-AUG-1996; 96WO-US13482.
XX	
PR	15-SEP-1995; 95US-0003784.
XX	
PA	(BAYU ) BAYLOR COLLEGE MEDICINE.
PI	O'Malley BO, Onate S, Tsai M, Tsai SY;
DR	MP1: 1997-202233/18.
XX	N-PsDB; AAT84543.
PT	Steroid receptor coactivator-1 polypeptide and DNA - regulated using
XX	a molecular switch, used in gene therapy
PS	Claim 3; Fig 1; 116pp; English.
XX	
CC	This polypeptide comprises human steroid receptor coactivator-1
CC	(SRC-1), a novel protein required for human progesterone receptor
CC	(hpr) transcriptional function. SRC-1 acts as a coactivator for
CC	hpr by reversing receptor sequeching. It is a coactivator for
CC	multiple steroid receptor superfamily members. The truncated
CC	C-terminal region of SRC-1 acts as a dominant negative regulator
CC	of steroid receptor function. An SRC-1 related disease can be
CC	treated by introducing SRC-1 nucleic acid (see AAT84543) into a host
CC	cell and infusing the cells into the patient causing an increase in
CC	the transcription of SRC-1 (claimed). A molecular switch can be
CC	used to regulate expression of a nucleic acid cassette incorporating
CC	an SRC-1 coding region for use in gene therapy. Transcription of a
CC	target gene can be decreased by providing a nucleic acid encoding a
CC	dominant-negative inhibitor of a SRC-1 polypeptide in a cell
CC	containing the target gene (also claimed).
XX	
SQ	Sequence 1061 AA;
<hr/>	
Query Match	13.5%; Score 1033.5; DB 18; Length 1061;
Best Local Similarity	28.8%; Pred. NO.1.3e-58;
Matches 348; Conservative 175; Mismatches 335; Indels 349; Gaps 52;	
QY	395 LNPISNSPFAHQAALCSGN--PGQDMTLSSNINFPINGKEQ-----MGMPKGRF 441
DB	10 vnp--slspahvavrstltpnsnmvstrln-----rfgssdlhsesshnsnsgqef 61
QY	442 GGSGGGMNHSHGV-----OATTPGGSNVYALKMNSPQGSSSPGMPPGTSMLSRHRHSPPV 496
DB	62 gcspsgqvlaavnalnkgqass-qskspstlnlnmpmeqtglslaq---fmsprtytstl 117
QY	497 ACSPRIPPEQFSR-AGSLSPSYGVCS--TGNShSYTNLSNALMALALSGHVSLASSIA 553
DB	118 atcrpmrnpsfpnnltstlsprygmesaeannmrrysnlpyvslsgmqemgnmnygfass 177
QY	554 SPDLKMGINIQNSPVNNPPPLSKMGSLSKDCFGLYGEPSEGGTTGOAESCHHPGEOKET- 612
DB	178 spvlrqmsqnsprsrlnlpg-akaekndkelaatlnemt-----qdadsddgxpkldeg 231
QY	613 ----NDPNLPNAVVSFEADQSRlhbskGCTKLQLLTYSDD-QMESPPLASSLSD----- 663
DB	232 llnhdnrl-----sdgdsksy--sqtbhkivqltltaeqqlrhaddctscdylact 281
QY	664 --TNKDSTGLSGSGSTHGTSLSKEKHILHRLDSSSPVDLAKTAAEATGDLSQESSS 721
DB	282 gtsnaasansgsqscpsnshltarkllmrllge-gpspdlttltsvepdkkd-sastev 339
QY	722 TAPG--SEVTITKOEVSFKKE--NALRLRLDKD-----TYDIGLETPTPKLERID 770

[illegible]

KW melanoma; PAS region.  
 XX  
 OS Homo sapiens.  
 XX  
 PN W09857982-A2.  
 XX  
 PD 23-DEC-1998.  
 XX  
 PF 17-JUN-1998; 98WO-US12689.  
 XX  
 PR 17-JUN-1997; 97US-0049728.  
 XX  
 PA (USGO ) US GOVERNMENT.  
 XX  
 PI Meltzer P, Trent JM;  
 XX  
 DR WPI: 1999-080946/07.  
 XX  
 PT New isolated steroid receptor co-activator, AIB1 - used to develop  
 PT products for the diagnosis and treatment of steroid-responsive  
 PT tumours, e.g. breast, lung, prostate or colon cancers or melanomas  
 XX  
 PS Claim 4; Page 30; 57pp; English.  
 XX  
 CC The AIB1 protein is a member of the steroid receptor coactivator-1  
 CC (SRC-1) family of nuclear receptor co-activators that interact with  
 CC estrogen receptors (ER) to enhance ER-dependent transcription.  
 CC The AIB1 gene is amplified and over-expressed in certain cancers  
 CC in particular breast cancer and steroid hormone responsive cancers.  
 CC The AIB1 polypeptide can be used to identify compounds which inhibit  
 CC ER-dependent transcription. Increased expression of the AIB1 gene  
 CC indicates aberrantly proliferating cells, thus detection of  
 CC increased expression of the AIB1 gene or an increase in the number  
 CC of copies of the AIB1 gene can be used to diagnose cancer or a  
 CC predisposition towards developing cancer. Compounds which inhibit  
 CC expression of AIB1 or compounds which inhibit interaction of AIB1  
 CC with steroid receptors or nuclear co-factors can be used for  
 CC reducing the proliferation of cancer cells. The PAS domain of the  
 CC AIB protein is a highly conserved domain among the SRC-1 family of  
 CC proteins. The PAS region functions as a protein interaction domain  
 CC which mediates binding between AIB1 and other proteins.  
 XX  
 SQ Sequence 186 AA;  
 XX  
 Query Match 7.9%; Score 604.5; DB 20; Length 186;  
 Best Local Similarity 60.4%; Pred. No. 1e-31;  
 Matches 113; Conservative 36; Mismatches 37; Indels 1; Gaps 1;  
 XX  
 QY 119 MLEALDGFEEFVNLEGNVFESENVTOYLRYNOEELMANKSVYSILHVGDDHTEFVNKLLPK 178  
 Db 1 lllgldgflfivnrdgnlvtfsenvtylqylgedlvtvsynlllheedkdlfkn-lpk 59  
 QY 179 SYVNGSGWSEGPFRNRSHTEFCRMLVPLPDESEEGHDNOAHOKYETMOCFAVSOPKSI 238  
 Db 60 stvngswetnehtgkqkhtfncrmllmktphlledlnaspmrqrgetmgdfalsqgram 119  
 QY 239 KESEEDLOSLICVARRVPMKERVPLPSSEFTRODQKITSLDSTMTMAAKPQMED 298  
 Db 120 meeedlqscmlcvarrittgerltfnspestlfrhdlsqkvndlnstsrmtprgfd 179  
 QY 299 LVRRCIQ 305  
 Db 180 lrrrciq 186  
 XX  
 RESULT 12  
 ABB60081  
 ID ABB60081 standard; Protein: 1778 AA.  
 XX  
 AC ABB60081;  
 XX  
 DT 26-MAR-2002 (first entry)

XX  
 DE Drosophila melanogaster polypeptide SEQ ID NO 7035.  
 XX  
 KW Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical.  
 XX  
 OS Drosophila melanogaster.  
 XX  
 PN W0200171042-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 PF 23-MAR-2001; 2001WO-US09231.  
 XX  
 PR 23-MAR-2000; 2000US-191637P.  
 PR 11-JUL-2000; 2000US-0614150.  
 XX  
 PA (PEKE ) PE CORP NY.  
 XX  
 PI Venter JC, Adams M, Li PWD, Myers EW;  
 XX  
 DR WPI: 2001-656860/75.  
 DR N-PSDB; ABL04184.  
 XX  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -  
 XX  
 PS Disclosure; SEQ ID NO 7035; 21pp + Sequence Listing; English.  
 XX  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins  
 CC (ABR57737-ABR72072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pcl\_sequences.  
 CC  
 XX  
 SQ Sequence 1778 AA;  
 XX  
 Query Match 5.6%; Score 425; DB 22; Length 1778;  
 Best Local Similarity 19.8%; Pred. No. 1.9e-18;  
 Matches 332; Conservative 197; Mismatches 583; Indels 566; Gaps 75;  
 XX  
 QY 93 NIDVOKSDVSTGQVIND-ALGPM-----MLEALDGF-----FVVNLEGNVFEV 139  
 Db 73 slhpvggvevsstpepllpelllqgvpeisayfealehyisgvgwllqvnangllsc 132  
 QY 140 SEANTOYLRYNOEELMANKSVYSILHVGDDHTEFVNKLLPKSIVN----- 182  
 Db 133 tqntrdllygkqglqlyqplymlyysgdhkl-----epintmynpnngnsans 186  
 QY 183 -----GSGWSGEPFRNRSHTEFCRMLVK 205  
 Db 187 gpggsasagtsagvgdlleelnngnsgsnsgagglgaggaagkkrstskvrm1vk 246  
 QY 206 PLPDESE-----EEGHNOEAHO-KYETMOCFAVSOPKSIKESEEDLOSLICVARRVPM 258  
 Db 247 dtrratgtssnceekplrrgshgdkyeavlliaa-----pvkdda-dassvllclitr-pe 300  
 QY 259 KERP-----LPSESEFTTRODQKITSLDSTMTMAAKPQMEDLVRRCIQKFA 309  
 Db 301 despleinlgqhvqgplleqmtfkldlhgklltdlpalrephqhtqwtvgrlqqlch 360  
 QY 310 QHEGESVYAKRHHHEVLROGLA-----FSQITRPSL-SDGTLVYAAQTKSLIR 357  
 Db 361 phd---lslkshlrldqdsasanspgagagatvsvsrfplrllgpadvvyhvXansrfl 417

Oy	358	SOI.NEPOLYSLHLMHREON-----VCVMNNDLT-----	387
Dd	418	nqcpgegdflmsvcllnseidmssnlgagsglglgiclamapslaasllsmd	477
Oy	388	----GOTMGKRLINISSNS-PAH--OALSGSN---GOMTJLSSNIHF-----INGP-	430
Dd	478	qlhggtsgassspaaagmlpthlllglyvggggggggnstqtetvsgrlmesalngty	537
Oy	431	KEOGMPIRGREGSGGNH-----VSGMOAT-----TPOGSNYALK-	466
Dd	538	lqqqqqqqrzgsassassanaalyafaspapaeahsfygdtfeidaahssfeldp	597
Oy	467	---WNSPSQSPGNMPCOPTSMLSPR---HRMSPEVASPRLIPSQPS-PHGSHSPVG	518
Dd	558	srgvgawdstprnasvaavptvtrppshgfspsvacspatpqylshsaasalpsqs	657
Oy	519	VCSSTGNHSHTNSLNALNALQELSEGHVSIGSSLASPDJMKCNLONSPVNMNPPLSKWG	578
Dd	658	nasgggnvygfnh-----sfdsadvpekdvdvqqnqqgsm-----n	697
Oy	579	SJDSKDFGLYGEPSEGTGOAESSCHPGOKETPNPMLPPAVSERADGOSRLHDSKGQ	638
Dd	698	sssnppllg-gglpngvggmlls-qqqqqqqt-----ppqqqqqqqqes-----	740
Oy	639	TKLQLLTTKDKDOMEPPPLASSL-----SDTNKDSTGSLGSSSTGTSTL-	683
Dd	741	serlrhlhtkqsn-----agglyglgdekyfkrpgaseekhaasgfkmgypqgmmmf	795
Oy	684	-----KEKHKLHLRLLODSSSPVDLAKTALFAETAGDLQSSESSPTPGSEVTIKOBP	734
Dd	796	gpmgsamrgyvgnasmmlhk-agnsqgnpm-lklilneks-edddngngsgggps-----	844
Oy	735	VSPKKENALLRYLLDKD-----DTKDIGLPETPKLERLDSKTDPASNT	779
Dd	845	-mnursgelmrqlknpdgsgshgmhrnaeqnmstedl-----kamklqsdpslnr	895
Oy	780	KLIANKTKEEKMSFEPPDOGSSELNDNEFLDLQNSLPOLPEPDRPCAPAGSVDKAI	839
Dd	896	k-----rslnepddopsakrs-----edkp-----sklct	920
Oy	840	INDLMQLTAENSPTYVGAQKTALRIISOTFFNNRPBGOLRLLPNONLPLDTLQSPGA	899
Dd	921	qkmaklakllqppeklypa-----pnpeqqlgyk	948
Oy	900	GPEPPIRNSSPYSVIPOPGMKNNGMIGNOGLNGSSTGICNGSASRPTMPGSEMAPOSS	959
Dd	949	-tlpoltitsvtscilaapnlisag-----stg-----pkaa	979
Oy	960	AVRTYCATTSAMNRPFVOGGMRIRPAASIMRRSSQPGORQTLOSVMYMIIGSELEMMNG	1019
Dd	980	anr-----nrkqqa-----qqqqqqqqqqqqqvagi-psq-----	1008
Oy	1020	GPOYSOOQAPAPNQCAPWPESILPIDQASFASQNROPFGSSPDDLICPHRAAESPEDCAL	1079
Dd	1009	-----qqqqqndvylsqggqqqlqspqlafqfnqqlattatsitlaastaasaaaaai	1066
Oy	1080	LDOYLIALRNFDELEIDRALG-----IPELYSOQOAVDPPEOFSSODSENMLE	1127
Dd	1064	lg-----egdselskllidswymeyppddrlivnapeseaalndiq-----kslmd	1109
Oy	1128	OKAPVPFOQVASQOMAMAGSTSPMQDPNFTFMKOQRSYATLRMPRPGRLRPGLVGYO--NQ	1188
Dd	1110	vesaafgndlnqqlmntq-----qqghqggqqqqqlalaqlaqggqqqrqhllqppay	1163
Oy	1186	PNOURL-OLOHRLQAQONROPLMNQJSNVSN-----YNLTLRLRG-	1223
Dd	1164	pqmllmqqqqqqqqqqqqlinqrlleamrngangfqrpmpmyparqfgymnavacpbg	1223
Oy	1224	--VFTOAIT-----NAQMIAQOREIELNHQHLPORONHOQOYOQFRLMARGCLMTMFSM	1276
Dd	1224	vvlpaqgqlnnlryqqqlaaaqke-----lljqqqqkqqllypenatigma-glmmigsj	1278
Oy	1277	VAPSMP-ATMSNBRIDQANAOGFPPBNYG-----ISOQDPGEFTGATPPO-----	1322

```

Db 1279 Inttqapnvsalstrlnp-sdaq-----lspnfaqltmgqqlspgrsreppspqpnqgyaqpf 1334
Oy 1333 -----SPLMSRAHATOSPMMQOSO-----ANPAYQAPSDING- 1355
Db 1335 qpqrqlspqqrqqqlsqqqqnnvqqqqlayqqqvgvdqgrsntpfrfgsnsgmgspsgmqnspp 1354
Oy 1336 -----MAOGNMGNS-----MFSQOSPFRHGQOANTSMY-----SNMNMINTSMA 1395
Db 1335 qwsqsgsgsgsgppgplpsgnaqrltqlqhnpmliaqlqvsyynaayqyqnnqtrqlnsppa 1454
Oy 1336 TMTG-----MSSNMQOMGOISMTSMYSVTSRSLGSMGPQOVNDPALRGNGLEPNQ 1446
Db 1455 vqpgpnpaqaqaalqrgnsfqqg-ggggqattpdsgsvgyfqpq-----spylgnvntvqq 1508

```

## RESULT 13

ID AAW78487 standard; Protein; 399 AA.

AC AAW78487;

DT 21-MAY-1999 (first entry)

DE Human SRC-1 ligand binding region.

KW Human; SRC-1; ligand binding; fusion protein; screening;

XX

XX TP11032767-A  
PN

09-FFB-1000

XX	16-JUL-1997.	97JP-0191007
PF		

XX 16-JUL-1997: 97JP-0191007.  
PR

XX  
PA (YAMA ) YAMANOUCHI PHARM CO LTD.

AA  
DR  
WPI: 1999-183824/16.

DN N-ESDB, WMAZ1304.  
XX

PT intranuclear receptor-combining protein

PS Claim 3; Page 10-11; 23pp; Japanese

CC This sequence represents the human SRC-1 protein ligand binding region

CC protein can be fused to reporter proteins and used for screening for

CC intranuclear receptor protein that bind the ligand.

SD Sequence 399 AA;

## Query Match

Matches 142; Conservative 70; Mismatches 139; Indels 84; Gaps 23;

QY 395 LNPISNSPAHQALCSGN--PGQDMTSSNINFPINGPKEQ-----MGMPMGRF 441

Db 7 vnp--sispahgvarsstlppnsnmvtrln-----rqgssdlhssshnsnsnsgsf 58

QY 442 GSGGMNHVSGM-----QATTPQGSNYALKMNSPSSPGMNPQPTSMLSPRHRMSPGV 496

Db 59 gcspsqivanvalnkqgass-qsskpslnlnppmegtglslag---fmsprqvtsgl 114

QY 497 AGSPRIPPSQFSP-AGSLHSPVCVCS--TGNSHSYTNSSLNALQALSEGHGVSIGSSLA 553

Db 115 atrprmpnsfpnnistlsspvgmtssacnnnrnsynlpvtlqgmegpnsvgfsas 174

QY 554 SPDLKMGILQNSPVMMNPPLSKMGSLDSCFGLYGEPSGEGTGOAESCHPGGEOKET- 612  
11 : : 1111 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : :  
Db 175 spvltqmsqnsprlniqp-akeaskdnkeastlnemi-----qsdnssdgypldsq 228  
QY 613 ---NDPNLPRAVSSERAGOSRLHDSKGTKLLOLTTRKSD-QMERSPLASSLP----- 663  
11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : :  
Db 229 llnhdrrl-----sdgdkry--sqtsbkivqllttaeqqltnadldtsckdvlsc 278  
QY 664 --TNKDSITGSLPGSGSTGSLKEKHKILHRLLODSSPVLDLAKTLTAETAGKDLQESS 721  
11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : :  
Db 279 qtsnasasanssgscpssthsitarkhlhrlldge-gpsdltltisvepdkkd-sastcv 336  
QY 722 TAPG---SEVTIKQEPSPVKKE---NALRYLDRDPTKIDGLPEITPKLERLDSKTD 775  
11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : :  
Db 337 svtqvgvqnsikleldaskkkeskdqllryllakde-kdl---rstpnlaldvk--- 389  
QY 776 ASNTKLIAMKTEKEE 790  
11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : :  
Db 390 -----vkvekke 396

## RESULT 14

ABB65772  
ID ABB65772 standard; Protein: 5533 AA.

AC ABB65772;

DT 26-MAR-2002 (first entry)

XX Drosophila melanogaster polypeptide SEQ ID NO 24108.

DE Drosophila melanogaster polypeptide SEQ ID NO 24108.

KW Drosophila: developmental biology; cell signalling; insecticide;

XX Drosophila melanogaster.

OS WO200171042-A2.

PN 27-SEP-2001.

PD 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2001; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX (PEKE ) PE CORP NY.

PI Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

DR N-PSDB; ABL09875.

XX New isolated nucleic acid detection reagent for detecting 1000 or more

PT genes from Drosophila and for elucidating cell signalling and cell-cell

XX interactions -

PS Disclosure; SEQ ID NO 24108; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent

CC capable of detecting 1000 or more genes from Drosophila. The invention is

CC useful in developmental biology and in elucidating cell signalling and

CC cell-cell interactions in higher eukaryotes for the development of

CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (AB116176-AB130511), expressed DNA

CC sequences (AB101840-AB116175) and the encoded proteins

CC (ABB57737-ABB72072).

CC The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPRO

CC at ftp.wipro.int/pub/published\_pcl\_sequences.

XX Sequence 5533 AA;

Query Match 4.4%; Score 334; DB 22; Length 5533;  
Best Local Similarity 20.3%; Pred. No. 1e-11;  
Matches 331; Conservative 197; Mismatches 654; Indels 450; Gaps 74;

QY 8 TSDPSRAETFRKKKECPDOLGSPKRNTERKRNDEKNKIIEELAEILIFANFDIDNFKP 67  
11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : :  
Db 2819 tsaptaqtiskry-----edkmetfifjlsdeesqf-p 2851  
QY 68 DKCALKEKVEYOIROIKOEKAAANIDEYOKSDVSSPTGGVYIDDALGPMLEALDQFF 127  
11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : :  
Db 2852 ega---etnk-----dlpsvstl-----gpivsaalq--- 2877  
QY 128 FVNNEGVNVEVSENVTOYLRYNOELMNKSVYSILHY--GDHTEFVKNLIPKSIYNGS 185  
11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : :  
Db 2878 -----tykgepsrpskneeahqqlvtchepegqqlerlsysgs 2917  
QY 186 WSGPPRRNSHTFNCRLVLPDSEEGHNDQEAHQKETYMQCPAVSQPSTI--KEEGE 243  
11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : :  
Db 2918 sssshadrerhrrekr--ekkrksqreqqg-lhqkskvet-kvdddnsvmddeagr 2973  
QY 244 DLQSLICVARRVPMKERVLPSSSEFTTRDLOGKITSLDSTYRAMKPMQMEDLVYRC 303  
11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : :  
Db 2974 aleaql-----msdltkrlpise-----ealprstaalyrtdmtd----- 3006  
QY 304 IQKFHAHQEGESVSYAKRHHHEVLROGLAFSQIYRFSLSDTGLVAAGTKSKLIRSQTNE 363  
11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : :  
Db 3007 vrfidnednsv-----dmktgyvkseqqeqhskd-----kkskkskskeeq 3051  
QY 364 POLVITSLHMLHREQNVYVNDPLTGOTMGK-----PLNDRSSNPAHQALC 409  
11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : :  
Db 3052 ek1---lqqrreslprnvastestapptpgklitvnvaaskhadqlldakhlhisp---pvc 3105  
QY 410 SGNPGQDMTSSNINFPINGREKMGMPMGREGSGGNHVSGMQATTPOGSNTVLAKN 469  
11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : :  
Db 3106 kpspslpcldqdddaalhtpaxktpksr--gndgl-----t 3142  
QY 470 PSQSSPGMNPQPTSMLSPRHRMSPGAVSPRIPSPQSPAGSLSPYGVCSSTGNSHY 529  
11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : :  
Db 3143 perekrp-----rlisp-----lptkpliansstlsgsaetp-----sgtvlvs- 3182  
QY 530 TNSSLNALQALSEGHVSLGSSLASPDLMKGNLQNSPVMMNPPLSKMGSLDSCDF--G 587  
11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : :  
Db 3183 -ssalattptestaags-----aap-----gldnspc-----saaqckkkesflpg 3224  
QY 588 LYGEPSGEGTGOAESCHPGGEOKET-----NDPNLPRAVSSERAGOSRLHDSKGT 640  
11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : :  
Db 3225 fdggldrtisesavgsi-saetnstslladnadepkhpvaaprratpdklleeksrvt 3283  
QY 641 LLQILTTKSDQMEPSPPLASSLSDTNKSDTGSILPGSGSTHGTSLKEKHKILHRLLODSSSP 700  
11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : :  
Db 3284 isq-----etesavsaallgesftgsttdyldgmdemsvneletrpl---viaepdee 3336  
QY 701 VDLAKTLTAETAGKDIS-----QESSSTAPGSEVTIKQEPV---SPKKENALLRY 747  
11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : :  
Db 3337 aalaakaletagepaslllepeemereaeapdpdeaeiesepvevaidpeelnkav--q 3394  
QY 748 LLDKDDTDIGLPEITPKLER--LDSKTDPASNTKLIAMKTERKEMFEEGDDGSGSLD 804  
11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : :  
Db 3395 slknedmndlkxad--ltpserdlqldtleenpedadsspgslkidelvsgssspkxsis 3452  
QY 805 NLEETILDLONSQLPOLPPDRPGAPAGSVQKQATINIMOLUTENSVVT--PVGAOKPAL 863  
11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : :  
Db 3453 nmsprrtandldpvn--eeqpklsnesttpgsvitklpfl---dtpktvpaglpesvp 3507  
QY 864 RISOSTFNN-----PRPGQIGR-LLPNONLPLD----- 890  
11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : :  
Db 3508 kieptlisklqpllvqpvqtlvpaphstgsgisansvlnldlsvniscntsaaasatas 3567  
QY 891 ----ITLOSPPGAGFPPIRRNSPYSVTPQPGMGNQGMIGNQNLGNSSTGCMGNNSR 946  
11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : : 11 : :  
Db 3568 asastisfgspstas-----qnamppastpckgpltpqgaaitrgslimpprlisipeqtp 3621  
QY 947 PTMPSGEWAPQSAVRVTCATTSAMNRPVQGM-----IRNPASISIPRPSQGGOROTL 1002





```

Db 3284 isg-----eetasavallgesfgtstldyldmdemsavneleptl---vinaepdee 3336
QY 701 VDLAKLAEATGKDL-----OESSAPGSEVYTIKOEPPY-----SPKKKNALLRY 747
Db 3337 aalaakateagepastlleepemereaeapdpdeeelesepvevildpelakav--q 3334
QY 748 LLDKDPKIDIGLEPIETPKLER---LDSKTPASNTKLIAKTEKEEMSFEFGDPGSELD 804
Db 3395 slkheedmmidkade--tpqserdlqldtdeenpeadeassgslikidetvgssspeksis 3452
QY 805 NLEETLDDLONSQLPOLFPPTRECAPAGSVDKAIINDLMOUTRENSPVT--PVAQAKTAL 863
Db 3453 nnsptetelanidipnv--esqpklsneslcpqpsvilkplf---dlpkltpvaglppspv 3507
QY 864 RISGSTFNPN-----PRPGQLGR-LLPNQLPLD----- 890
Db 3508 kileptlisklqgplvqpvqetvlpephstgsgisavsvinldlsvlscntsasatas 3567
QY 891 -----ITLQSPGACFPPIRNSPPYSVIPOPAMNGMIGMGNLGNSSGTGMGNASR 946
Db 3568 asasisfsgptas-----qnamqgastpkqgpltpqgalrtqslmqpplispeqtph 3621
QY 947 PTMPSGEMAPQSSAIVRTCAATTSAMNRPVOGGM-----IRNPAASIPMRSSPGQKQTL 1002
Db 3622 fawpqnvlspqs-----hhpqgpltymwgjitaps---phspilhsprgvaq 3664
QY 1003 OSQVMNIGPSELEMMNGGPOYS-----OOQAPPNQ-----TAPPESTLPTDAQS--FAS 1050
Db 3665 srlvgqisp-----vgipmwsgpspqgqvqqlqghalltspsgsnlspiaspltrvlis 3718
QY 1051 QNRQPFCS-----SPDLLLCP-HPAESPSEDEGALLDQVLYLALNPFGLLEIEDALGIPE 1104
Db 3719 ssnspstksnysyprnqrpqgspksvaevqtlmtcltpikmtpidqgphn-----pt 3774
QY 1105 LVSGSOAVDEPQESSQD-----SNML-----EQKAPVPPQOYASQAQMAQG 1146
Db 3775 llskvelevpqgqatqsgvasspplgsjlpnhkvnlnahbqgqgqvavlakmabqhqmq 3834
QY 1147 SY-----SPMODNFMHMGORPSVATLRMQPRPGLARTGLVONP---NQLRL--- 1191
Db 3835 qfimbqgmrlqrgqhmqqqlhgssqqltsa---pqhmqhqbhaqqgqghnqhnlngq 3889
QY 1192 -----OLOHRLQAOQNRQPLMNOISNVSNNLTLRPVTPQAPINAOQMLAQOREIL 1243
Db 3890 lhaqghlctqkqhqqgqfngqlqghqsgq-----qhqyqgqgqgqghnsg 3935
QY 1244 NQHLRQRMHQOQVOQRTL-----MMNG-QGLNMTPSKVAAPSG-----MPATMSNPR 1290
Db 3936 qqbhsqgqlnqbhaqqgqlqgkqlqgmhbpqgqkspgqvghlsgstslfasqghnsg 3995
QY 1291 IP-----QANAOOF-----PPRPVNGIS-----OQPDGFGATTPQSPPLMSPRMAHQOS 1335
Db 3996 lpargvppqgqhpqqlsthspeckpntlvsngvgvppalllrvsgshsqnqgqqlphqgqs 4055
QY 1336 -----PMOQSOQ---ANPAYQAPSDINGMAQGMGNSMFSQOSPHFGQOATSMYSNNMN 1389
Db 4056 sgphbqgqlaspganlrlqplnvrlqntpkllvqghlvaqnvpppqgqnalhyppqgq 4115
QY 1390 INVSMAINTGMSMNMQTCGISMTSV-----TSVSTGSLSMGEQVNDPLRGCNLFPN 1445
Db 4116 kd-----stppghvepctpsmsaqktsesvsvlrltpcltqlavlsantvgs--lltenlilki 4171
QY 1446 QLPQMD-MIKOE 1456
Db 4172 sqpkqdelieqd 4183

```

```

DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 11742.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical.
XX Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE ) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW,
XX
DR WPI; 2001-656860/75.
DR N-PSDB; ABL05753.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Disclosure; SEQ ID NO 11742; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL10511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABBS7737-ABBS2072).
CC
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcl_sequences.
XX
SQ Sequence 2280 AA;

```

Query Match 4.48; Score 332; DB 22; Length 2280;  
Best Local Similarity 20.68; Pred. No. 3.5e-12;  
Matches 266; Conservative 138; Mismatches 452; Indels 438; Gaps 59;

```

QY 426 PINPKEDQMGMGRFGSGGMNHS-----GMOATTPQG-----SNY 463
Db 1017 plngdaad-gasvvgigqagvgvmmnasgplisagqqlatqpsllnrvtqlnvqlstk 1075
QY 464 ALKMNSSQSSPGM--NPGQPTSM-----LSPRHRSBGVAG-----SP 500
Db 1076 kctasapalsprvlllepeqdcskkykhanglryhshabhaqgagsamedmsqapedp 1135
QY 501 RIPPQSPAGSLSPVGVCSSTGNSHSYTNSSIALQALSEGKGVSLGSLASLPDLKMG 560
Db 1136 atppsp-----gvasgtsgsasvasavpa--lagaqg---gtvaavspnplla 1179
QY 561 NLONSVPVNMN-----PPPLSKMGSLSKDCFCGLY 589
Db 1180 nsnrvptnngnavapsapatgsvtiaapntpsvvetqapltpppyt----- 1225
QY 590 GEPESEGTGAESSCHPEQKETND---PNLPPAVSSERADGQSLHDSKQTKLL--- 642
Db 1226 --ppapricavav--pqaegsvsvlplgnlpltagpnastqgqpplqgqpllvppq 1281
QY 643 -QLITTKSDQMEPSPLA-----SSLSDTKKSDTGSILPGSGSTRHGTSLAKKKHILH 691
Db 1282 gsaaslgqgqgqgqpvavagsgltaglsqgalsqhqqlmgqilp-----amlsdqgq-q 1332

```



QY 303 CIOGFHAESESVAKRHHHEVLROGLAFSQRFF----- 339  
 Db 353 sl--fscqgads-erlmatfkyslkgqgetsyrrfigkygycwllsqatlyvdklkr 409  
 QY 340 -----SLSSGTLVAOTKSLIRKSLQSTNFPQVLYSLHMLHREONV 379  
 Db 410 qsvvcynvynslnlnkheiyiaqtaaseqkqhgaeekepekaadpelliaqetke 469  
 QY 380 CVMNPDLTGOTMGRPLN-----PISSNSPAHQ-AICSGNPGQDM 417  
 Db 470 tvnprhntselqakpqlqeseakaektieektatiiprvatastaadqiklipesnpykqi 529  
 QY 418 TLSSNINFPINGPREQMGMPGREGSGGMNHYSGMOATTPQSGSNYALKMNSPQSOPGM 477  
 Db 530 l-----gaellikr-----enhspgrlitaq-----lsgssqgl 560  
 QY 478 NPGQ--PTSMLSPRHRMSPGVAGSPRIPPSQSPAGSL--HSPVGVCSSTGNSHSYTSS 533  
 Db 561 rpeekrkksvtaavlylrpsp---appllrp---plavlccktlpiv----- 599  
 QY 534 LNALQALSEGHVSLGSSIASPDLKMGMLQNSPVNMNPNPLSKKGSLSKDCFGLYGEPS 593  
 Db 600 -----epnl-----ppltaataais----- 615  
 QY 594 EGTGQAESCHPEQOKETNDPNLPRAYSSERADQGRUL--HDSKQOTKLQTLTTRKSDQ 651  
 Db 616 -----ssngqliaqgcqlqppqp---aqdmksyfcslfaddgrglmlke----- 659  
 QY 652 MEPSPLASSLSDTN-----KDSYGLPSGSGTHGTSLSKHKHKLHMLLDDSSPV 701  
 Db 660 -epddlsbhlastnciqlidemtfisdmlvglmgt-----cllpedinsl 702  
 QY 702 DIAKLTAEATGKDLQSSSS--TAP-----GSEVTRKQEPVSPKKENALLKYLKDDTK 755  
 Db 703 ddtcstaagqhyqspssstsepsntssmsyansplsp----- 744  
 QY 756 DIGLEITPKLERLDSFTDPASNTKLIAKTEKEMSFEPDQPSSELDLEILDDLON 815  
 Db 745 -----ltp-----nataasn-----pshqgqghnqgq-----qg 771  
 QY 816 SOLPQLFPDTRPGAPACSVKQAIINDLMQLTANSPTVPGAKOTALRISQSTFNPNRP 875  
 Db 772 qqqqghnpqhdnsnss-----nidplfnyeesndt-----scsqhlhpsltskspe 822  
 QY 876 GOLGRLPLNOMLPL--DITLOSPTGAGPFPIRNSSPYSVPOGMMNGMIGNQNLG 933  
 Db 823 ssplslspnslqgedffsfeaamrpylpidddmp--litedclm----- 867  
 QY 934 NSSTGMIGNSASRPTMPSGEMAPQSSAVRVYCAATTSAMNRPVOGMIINPASIPMRPS 993  
 Db 868 -----wcpedlqtmv----- 878  
 QY 994 SQRGQROTLOSQVWNITPSELKMMGGRQYSQOQAP-----NOTAPWPESTL----- 1041  
 Db 879 --pkeldalqgqlqgqghqyagnltyqgqgqgqqlqghfsmalscpastvasls 936  
 QY 1042 --PIDQASFAQNQPRGSSPDD--LLCPHPAESPSDEGAL----- 1079  
 Db 937 pspvqg-qhqqgqaavfstssetaallc-----gsngtllslilsgsvtyaeeener 988  
 QY 1080 -----LDOLYLALNFDLEIDRALGIPELVSQAVNDEQFSSQSDSNIMLEOKAFVP 1134  
 Db 989 lqghqgqgqgqtsqnefttfqqlqgeqlqgeeqqrgqqgqgqgqqlslisnieckk 1048  
 QY 1135 QOYASQAOAGSYSPQODNFHTMGORPSTATLRMQPREGLRPTGLVONO-----P 1186  
 Db 1049 eky--dvymgmslchpmed-----afendyskdsanldcwdlilqmvvdtepvsp 1096  
 QY 1187 N-----QLRLOLHRLQAOQNRPRLMNQISNVNLTLLAPGVPTQAPINAO 1234  
 Db 1097 naasptckvsaql-lqgqgqlqgqgqgq-----qnlilaavp-----ll 1136  
 QY 1235 LAOROREILNQHRLOROMHOQOQVOQRTL--MMRQOGL--NMTPSNVAPEGMPATMSNPR 1290

Db 1137 tiqnmkelmqgqgqgqgqgqqlqgpaikllngaslapvntkatiirlveskppttqsr 1196  
 QY 1291 IPQANAOQFPFPYVYISQODPG-----FTGATTPQSPILMSPRM-----AHTQSPMM-- 1338  
 Db 1197 makvnl-----vppgqghnkrhlnsaltaqgnpveskrlksqclcldvgsqqlq 1246  
 QY 1339 -----QOSQANPAYQAPSDINGMAQNMCGNSMFQSOQSPHRFGQOQANTSMYSNNMNIN 1391  
 Db 1247 qlqkdpeqgqqtgaakrage--rw-----qlsaesk-----qkqgqgqsnsvlkn 1291  
 QY 1392 VSMATNTGMSMNMQMTQGISMTSVTSVTSGL-----SSMGPPQVNDPALRGCNLF 1443  
 Db 1292 llvsgrdddeeamllidednslvgpplpkygylplohctsssvlrdyhmnpilisgnf- 1350  
 QY 1444 PNQLP-----GMDMIKQEGDT 1459  
 Db 1351 --qlspvfygsdssgqgdet 1368  
 RESULT 18  
 AAW79533  
 ID AAW79533 standard; Protein; 846 AA.  
 AC AAW79533;  
 DT 02-FEB-1999 (first entry)  
 XX  
 DE Human CLOCK protein.  
 KW CLOCK; circadian rhythm; human; jet-lag; sleep-wake disorder;  
 KW seasonal affective disorder; cancer; transcription factor;  
 KW diagnosis; therapy.  
 OS Homo sapiens.  
 XX  
 FH Key  
 FT Protein  
 FT /note= "Location/Qualifiers  
 FT Protein 2..846  
 FT /note= "Claimed fragment"  
 FT Protein 10..846  
 FT /note= "Claimed fragment"  
 FT Protein 11..846  
 FT /note= "Claimed fragment"  
 FT Domain 35..81  
 FT /label= bHLH  
 FT Domain /note= "basic helix-loop helix domain"  
 FT /label= PAS-A  
 FT Domain 273..318  
 FT /label= PAS-B  
 XX  
 PN W09840514-A1.  
 PD 17-SEP-1998.  
 XX  
 PE 13-MAR-1998; 98WO-US05114.  
 PR 30-JUN-1997; 97US-0885291.  
 PR 13-MAR-1997; 97US-0816693.  
 PA (NCUN ) UNIV NORTHWESTERN.  
 PI Pinto LH, Takahashi JS, Turek F;  
 DR WPI; 1998-520828/44.  
 DR N-PSDB; AAW61450.  
 XX Novel mouse and human circadian rhythm gene, clock - useful for  
 PT treating e.g. jet-lag, sleep-wake disorders, abnormal cell division,  
 PT etc  
 PS Claim 20; Fig 14.1-14.3; 154pp; English.









QY 1168 -----LRMPPRGLRPTGLVONOPNOLQLOHRL-----QAQ 1200  
Db 1508 spapnwgagpprfgaapppg-aphbpriqgpaqvagdqhcypppqgpppppqgqqgqq 1566  
QY 1201 QNRQPLAMQIN-----VSNQITLIRPGV-PTQAPINAOMLAROREITINQ 1245  
Db 1567 qqqqppqvagpppqgppqgppqagppqtaqsgiaappspilrppsgpqgqrmppgm 1626  
QY 1246 HLRRORHMQOQVOORFLMAMGQGLNMTPSAVAPSGMATSNPRLIQAN-----AQQ 1298  
Db 1627 paqgqsgqsgqgvpq-----pppqgashgvp-----spglpqvgpggmwkpypa 1671  
QY 1299 FPFPPNVCISQDPDGFATTPOSPLMSPRMAHTQSPMAMQSOA-----NPAYOAP 1350  
Db 1672 mppppsgqvgqvgqgpppgmmssqkppmpmqagmqqp1qgqppshqphbqhphqhp 1731  
QY 1351 SDI--NGMAGCGMCGNSM 1366  
Db 1732 hqmpnpqtaqpgyppgm 1749

## RESULT 22

AB67600  
ID AB67600 standard; Protein: 642 AA.

AC AB67600;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 29592.

KW Drosophila; developmental biology; cell signalling; insecticide;  
pharmaceutical.

OS Drosophila melanogaster.

PN W0200171042-A2.

PD 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US09231.

PR 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

PA (PEKE ) PE CORP NY.

PI Venter JC, Adams M, Li PWD, Myers EW;

DR WPI: 2001-656860/75.

DR N-PSDB; ABL11703.

PT New isolated nucleic acid detection reagent for detecting 1000 or more  
genes from Drosophila and for elucidating cell signalling and cell-cell  
interactions -

PS Disclosure; SEQ ID NO 29592; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent  
XX capable of detecting 1000 or more genes from Drosophila. The invention is  
XX useful in developmental biology and in elucidating cell signalling and  
XX cell-cell interactions in higher eukaryotes for the development of  
XX insecticides, therapeutics and pharmaceutical drugs. The invention  
XX discloses genomic DNA sequences (AB116176-AB130511), expressed DNA  
XX sequences (AB101840-AB116175) and the encoded proteins  
XX (AB57737-AB572072).

CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 642 AA;

Query Match 4.1%; Score 311.5; DB 22; Length 642;  
Best Local Similarity 23.0%; Pred. No. 1,1e-11;  
Matches 160; Conservative 86; Mismatches 273; Indels 177; Gaps 34;

QY 7 NTSDEPRAETRRKCCPCPOLGSPKRNTEKRNREQENKYEELAEILFANFNIDNENFK 66  
Db 5 nldkerfaste-nhc-----elerrirnmktaylclsdmvp-----ptcalark 49  
QY 67 PDKCILKEVTKQIQRQIEQEKAAANIDVOKSVSTG---QGVIDKDALGPMAL 123  
Db 50 pdklilrmavahmkalr-----gtqltsdgytkpsfllldqelkhlillea 96  
QY 124 DGFEEVNLK-GNVFESENVTOYLRYNOELMKNVSYSLHVGHTFEVKNLDPKSLVN 182  
Db 97 dgflfvscdsgvlyvdsvtvplnylqsdvgtsllyehlpddrekireqlscdesq 156  
QY 183 GGS---WSGEPFRNRSH-----FNCRLVPL-PDSEEGHNOEAHQ----- 222  
Db 157 agrlidlkgctvckeghqsgmslmqarrgflcrmrvgvnpesvsglnrlkqnslg 216  
QY 223 -----KRTMQC--FANSQPKS-----IKEGEDLS--CLICVAR-----KVPKER 261  
Db 217 psrdgtayavvchcglxlmwptldmfpmlmeardvdmshoclaigrlyvstcaadm 276  
QY 262 PVLPSSEFTTRDLOGKITSLDTSTMRAMKPGWEDLVRCIOKFKHQHEGESVYAKR 321  
Db 277 sgsmngsefllrhamdgyftfvdqrvlnllyltptellgkiclydfnp-----edslnke 332  
QY 322 HHHEVLRQ-GLAFQIYRFLSDGTLVAQTKSKLRSQTNEPOLVLSLMLHREQNV 380  
Db 333 sfdygkkgqgmfslllyaraknseywvrltqayaflnpyldeveyl-----vc 381  
QY 381 VMNPDLTGQTM-GKPLNITSNSP-----AHQALCSGNPGQDMTLSSNINP 430  
Db 382 t---nsqglmhgapldaaahlteqvqgqgqghvyvgaaydva----- 426  
QY 431 KEOMGMPMRFGSG-GNHNVSMOATTPGCSNVYALKNNSPQSSPQANNPCOPTS-MLSP 488  
Db 427 rrel-tpvgsatndmgymtamlamaptpq-----qqgqgqgqpgsaqgttptytydt 478  
QY 489 RHRMSPGVAGSP---RIPSQFSPAGSLHSPGVGCSSTGNSHSTNSLNL----- 537  
Db 479 th--spysaagppslaklpkqgstp-----cpv-----apnswaalrpgqqg 519  
QY 538 --QALSECHGVSLGSLASP-----DLKMN-----LDNSPYNMNPPPLSKMS 579  
Db 520 qqqpvttegyyqyqtsparspqptytqlsagngnrgqagpayaqgp-----pppnapgm 575  
QY 580 LDSKDCFLYGEPPSEGTGQAESSCHPGEQKETNDP 615  
Db 576 wdwgqagq-hphhphltahphhphhphhpgpagaqgp 610

## RESULT 23

AAW79529  
ID AAW79529 standard; Protein: 855 AA.

AC AAW79529;

DT 02-FEB-1999 (first entry)

DE Mouse CLOCK protein.

KW CLOCK; circadian rhythm; mouse; jet-lag; sleep-wake disorder;  
seasonal affective disorder; cancer; transcription factor;

KW diagnosis; therapy.

OS Mus sp.

XX Location/Qualifiers

FH Key 2..855  
FT Protein /note="Claimed fragment"  
FT Protein 10..855

```

FT      /note= "Claimed fragment"
FT      11..855
FT      Protein
FT      /note= "Claimed fragment"
FT      35..81
FT      Domain
FT      /label= bHLH
FT      /note= "basic helix-loop helix domain"
FT      115..163
FT      Domain
FT      /label= PAS-A
FT      273..318
FT      Domain
FT      /label= PAS-B
FT      514..564
FT      Peptide
FT      /note= "deleted in mutant CLOCK"
XX
XX      W09840514-A1.
XX      17-SEP-1998.
XX
XX      13-MAR-1998: 98WO-US05114.
XX
XX      30-JUN-1997: 97US-0885291.
XX      13-MAR-1997: 97US-0816693.
XX
XX      (NOUN ) UNITV NORTHWESTERN.
XX
XX      Pinto LH, Takahashi JS, Turek F,
XX      WPI: 1998-520828/44.
XX      N-PsDB; AAV61401.
XX
XX      Novel mouse and human circadian rhythm gene, clock - useful for
XX      treating e.g. jet-lag, sleep-wake disorders, abnormal cell division,
XX      etc
XX
XX      Claim 18; Fig 12; 154pp; English.
XX
XX      This is the amino acid sequence of mouse CLOCK protein, an integral
XX      component of the circadian clock that serves to regulate various
XX      aspects of circadian rhythm, including intrinsic circadian period
XX      and the persistence of circadian rhythmicity. The sequence was
XX      deduced from the isolated clock gene (see AAV61401). Human CLOCK (see
XX      AAW79533) is also claimed. CLOCK is a member of the bHLH-PAS domain
XX      family of proteins, and thus likely interacts directly with DNA.
XX      It has Glu-, Pro- and Ser-rich C-terminal regions characteristic of
XX      activation domain transcription factors. The invention provides
XX      isolated and purified CLOCK polypeptides, polynucleotides, vectors
XX      and host cells. The polynucleotides or polypeptides can be used to
XX      treat disorders of altered or disrupted circadian rhythms e.g.
XX      jet-lag, seasonal affective disorder, sleep-wake cycle disorders
XX      such as mood state, stress, neurological disorders, to regulate
XX      diet and food intake especially for diabetes, to treat
XX      cardiovascular, respiratory, liver or endocrine disorders, and for
XX      diagnosis and treatment of abnormal cell division such as cancer.
XX
XX      Sequence 855 AA:
XX
Query Match 4.1%; Score 309.5; DB 19; Length 855;
Best Local Similarity 21.7%; Pred. No. 2..2e-11;
Matches 150; Conservative 107; Mismatches 254; Indels 181; Gaps 27;

```

```

Db      205 kepstevevrrfignfklsfsvstctingfegltqrthrpseyedrvctvatvrlatpqlk 264
Oy      240 EEEGEDLQSCILICVARRRPMKRPVLPSPSEFTTRRODLQKITLDTSTMAAKPMEDLT 299
Db      265 e-----mctv-----eep-----neeftrshlskwkflfld---hnapllylpt 302
Oy      300 VRRICQFHAQHGESESVAKRRHHNEVLROGLAFSQIYRPSLSDGTLVAQTKSKILRSO 359
Db      303 evlgtsgdydyhvddlenlakcheh-lmgygkxgkscyrtfltkgqw'iwlqthyltlypq 361
Oy      360 TTNEPQLVISLHMLHREQNVCMNPDLTGOTMCKPLNPISSNPAPNALCQSGRPGDMTL 419
Db      362 wnsrpeflvcthtvsvyaetraerrelg-----leeslpetaadksqsgsdnrt 412
Oy      420 SSNINFPINPKRQMGMPKRFQSGGGMNHVSGQMATTPOGSSVYALKMNS-PEQSSPGMN 478
Db      413 nt-----vsikealerfdhs-----ptpsasrrsrkshtsvspst 451
Oy      479 PGQ-PTSMLSPP-----RHRMSPGVAGSPRIPPSQFSPAGSLHSPVGVCCST 523
Db      452 ptkipdtstpprqlhpbahemktrrsfsgslnsqvspsltcpamsqaanlpl'-pq 509
Oy      524 GNSHSTNSSLNMLQALSE-----GNGVSLGSSL 552
Db      510 gmsqfqlfqsaglgamqhlkdqlqgrtmieanlhrqgeelkldqqlgmvhgqglqmfllqg 569
Oy      553 ASPDLKMGNLQNS-----PVNMNMP--PLSKMSLDSKDCFGIYGEPESEG----- 595
Db      570 snpqlnfgsvqlsgnslqqltprvmgqgvypangvqs-----ghistqghmld 619
Oy      596 -----TTGQAESCHPGEQKFTNDPNLP 620
Db      620 qtlgslstqsgqsvmsqnsqgtslpsqfqs 651

RESULT 24
AAV32214
ID      AAV32214 standard; Protein; 855 AA.
XX
XX      AAV32214;
XX
XX      15-FEB-2000 (first entry)
XX
XX      Mouse CLOCK protein.
XX
XX      Clock; mouse; transcription factor; circadian rhythm;
XX      jet lag; sleep disorder; depression; seasonal affective disorder;
XX      fertility; therapy.
XX
XX      Mus musculus.
XX
XX      W09957137-A1.
XX
XX      11-NOV-1999.
XX
XX      06-MAY-1999: 99WO-US10072.
XX      07-MAY-1998: 98US-0084610.
XX
XX      (HARD ) HARVARD COLLEGE.
XX      PA      Weitz CJ, Gekakis N, Staknis D;
XX      PI      WPI: 2000-052938/04.
XX      DR      N-PsDB; AA234627.
XX
XX      Novel heterodimeric composition for identifying modulators used in
XX      PT      diagnosing and treating circadian clock disruption disorders -
XX      PS      Disclosure; Fig 16; 96pp; English.
XX
XX      This sequence represents murine CLOCK protein. CLOCK activates the

```



```

Db 281 iagvsakpavtapsglnhplg--hpllpvatsgsmniaisspllvnlq--ndgnqms 336
OY 548 LGSSLSAPDLMKMGNLQNSP-VNNMNP-----PPLSKMGSLSKDCFGLYGE-----PS 593
Db 337 nnpqlksppqgq---qgspligmspsaqgmmspmrscspatpsdf--lmgdvlspvyps 391
OY 594 EGTGGAESSCHPEOEKETNDPNLPPAVSSERADGOSRLHDSKGQTKLLQLTTKSDME 653
Db 392 spltpgaaa-----ppcpvymrnlqqgqqqf-----qgmals 425
OY 654 PSLASLSLSTNKDSTGSLPGSGSTHETSLKEKHKILHRLDSSSPVDLAKLAEATGK 713
Db 426 psqnaahlyfqgqg-----gppqqgqqgqgpphrfqgq-----461
OY 714 DLQSESSPTAPGSEVTRKQEPVSPKKKENALLRLLDKDQTKDIGLPEITPKLERLDSKT 773
Db 462 --hqmgsppq-----valrqq-----qldr-----479
OY 774 DPASNTKLAMKTEKEMSF-EPGDQPGSELDLNEELIDLQNSQLPOL---FPDTRPG 828
Db 480 -----qgqmrftmqppqg-----lqgqqgqqgqfqpapdcfn 511
OY 829 APAGSDKCAIINDLMQLTAEVSPVPGAOKTALRISOSTFNNRPBGQLGRLLPNQNL 888
Db 512 mngpmgqqqftrhqlrpgjrlappphpgqqqqrllmglsmpqasppq-----560
OY 889 LDTTSGPTGAGPF-PIRNSPSYVLPQPMGNOGMINOGNLGNSSTGMINSNAS-R 946
Db 561 -----qmsqasplgplspasshshspjmsnhdqgqqgqgq-----mvspsatpg 608
OY 947 PTMPS--GEMAPQSSAVRTCAATTSAMNRPVQGMIRNPASI--PMRPSQPGQRQT 1001
Db 609 ptapsemagqlmpanghvp---aapppdyngaavns--xrlaglnkpmatsatksrf 663
OY 1002 LQSQVM-NIPSELEMMMGPRQYSQQAAPNQTA-PMRESLIPDQASFAQONQPRESSP 1060
Db 664 lrygmynlqqqqqvsllpqpqqqqqqqqqlppr--alrtpq---qqqqsghvqp 717
OY 1061 D-----DLCPHFAAE--SPSDGALIDQLYLALRNFQDEIEDRALGIPELVSQS 1109
Db 718 qvsgqgqgqvdsllsivdtltnddallptlncldstlsiedkne---lesllqda 774
OY 1110 QAVD---PROFSSQDSNMLE--QKAPVPPQYASQAQMAQGSYSPMODPNFHTMGORP 1163
Db 775 kdldldiednlseavmrdaIntlgeapleqqgqqgqgqaa-----mqp 818
OY 1164 SYATLBMQRRPGLRPTLVONOPNQLRLQHLQAOQONRQPLMNQISNVNVLTRPG 1223
Db 819 qqqqtaqummp-----yqqpqqqqqqmqqlmmqqqrqqqlpq-----857
OY 1224 VPTQAPINAQMLAQRREILINQHLRORQHQQOQVOQRTLMNRQGLNMTPSMVAAPSGMP 1283
Db 858 -----mqgrqqqlqrqlqlqqqqqqqqqlqlvqqqqqqqqqlqy-----900
OY 1284 ATMSNPRIPQANAO-QFPPPNRTGISQDPDPTGATTPQSPMLSPMAHTQSPMQOSQ 1342
Db 901 ---qhpqgqqrppqkfllmptbg--dmepmasdsceaeaeleaqpfirnsbg--lgsfn 955
OY 1343 ANPAYQ-----APSDINGMAQGNMGNSMPSQSPPHFGQOANTSMNSNNINIVSM 1394
Db 956 fupandmlllptnlfseedsnsaqhpmg---lsedq-----ersdlslnsatsarar 1006
OY 1395 ATATGSGSNMOMTGOISMTSVTSVSTGSLSMGPEDVNPALRG--NLFPNQJPGM 1450
Db 1007 ktqvskanhnsfseagdnsprrssnspilglns--pgsvn--anaqgspkkaapnllrkg 1062
OY 1451 DMIKQEGDTRRK 1462
Db 1063 ---lqgqvkerk 1071

```

```

AAW37097
ID AAW37097 standard; Protein, 870 AA.
XX
XX AAW37097;
AC XX
XX 27-MAR-1998 (first entry)
DT XX
XX Human endothelial PAS domain protein 1 (EPAS1).
DE XX
XX Endothelial PAS domain protein 1; EPAS1; screening assay; gene therapy;
KW endothelial tissue specification; EPAS1 binding; heat shock protein;
KW basic helix-loop-helix motif; bHLH; hypoxia inducible factor;
KW HIF-1 alpha binding site; binding affinity; genetic hybridisation screen;
KW Immunogen.
XX
OS Homo sapiens.
XX
XX US5695963-A.
PN
XX 09-DEC-1997.
PD
XX
XX 17-JAN-1997; 97US-0785241.
PF
XX
XX 17-JAN-1997; 97US-0785241.
PR
XX
XX (TEXA ) UNTV TEXAS SYSTEM.
PA
XX
XX McKnight SL, Russell DW, Tian H;
PI
XX WPI, 1998-041300/04.
DR
XX N-PSDB; AAV00641.
DR
XX
XX Endothelial PAS domain protein 1 proteins - used for isolating EPAS1
XX regulators, e.g. a heat shock protein
XX
PS Claim 1; Columns 25-30; 23pp; English.
XX
XX The present sequence represents a novel human endothelial PAS domain
XX protein 1 (EPAS1). EPAS1 proteins can regulate specification of
XX endothelial tissue, such as vasculature, the blood-brain barrier, etc..
XX The protein can be used in a screening assay for agents that
XX modulate binding of EPAS1 to a binding target, especially a basic
XX helix-loop-helix (bHLH)/PAS protein, a heat shock protein or hypoxia
XX inducible factor (HIF-1 alpha) binding site. The EPAS1 protein is
XX incubated with the binding target and a test agent and the effect of the
XX test agent on the binding affinity of the protein for the target is
XX determined. The proteins may be produced recombinantly from transformed
XX host cells from the subject EPAS1 encoding nucleic acids or purified from
XX mammalian cells. The proteins may be used in diagnosis (e.g. genetic
XX hybridisation screens for EPAS1 transcripts), therapy (e.g. gene therapy
XX to modulate EPAS1 gene expression) and in the biopharmaceutical
XX industry (e.g. as immunogens, reagents for isolating B-cell specific
XX activators or other transcriptional regulators, reagents for screening
XX chemical libraries for lead pharmacological agents, etc.).
XX
SQ Sequence 870 AA:

```

Query Match 3.8%; Score 292; DB 19; Length 870;  
Best Local Similarity 20.6%; Pred. No. 3.2e-10;  
Matches 217; Conservative 188; Mismatches 406; Indels 294; Gaps 46;

```

OY 10 DPSRAETRRKRECPDQIGSPKRTNERNEQENKYIEELA-ELIFANFNIDNFNFKPD 68
Db 6 ekkrsserke-----ksrdarccrsktelevyelahelplph-----svssld 52
OY 69 KALIKETVQIKQIKOEAAAANIDVOKSDVSTGCGVITDKDALGPMMLDALDGF 128
Db 53 kasimrleisflrthklssvcseeseeaad-----qgnduliykalegfia 100
OY 129 VVNLEGVNVFVSNVVOYLAVYNOELMNKSVYSILIHGDHTEFYKNLSPRSYNGSGWSG 188
Db 101 vvqdgdmllfiseniskrmgltcqveltgnsifdthpcdheelelreml---slkngsgfg- 156

```

```

OY 189 EPPRRNSHTFNCRLMKLPDSEEGHNDQEAHQKYTEMOCFAVSGPKSKEEG----- 242
DB 157 -----kxskdmsterdfmfmkctvtgrgtvnlkstwtvyl 193
OY 243 -----EDLOSCLICVARRVPKKEPVLPR-SSESFTTRDLOGLK 279
DB 194 hctgqvkvynncpnhslcgykcpilscilncepilhmdipldskftlshmdmk 253
OY 280 ITSIDTSMRAMKRCW--EDLVRCIOKFNHQHGESVSYAKRHHEVLRGLAFSQT 337
DB 254 ftycd-----rtelilyhpeellgrsayefynaldsemnt--kshqnlctqgyvvsqy 307
OY 338 RFSLSIDGTVAQAQTSKILRSQTNPOLVLSLH--MLHREONVCMVMPDLTGOMGRP- 394
DB 308 rmlakhggyvyletqgtvilynpriqlpgqimcvnyviseiekndvfmtdt-estlfkph 366
OY 395 ---LNPISNSPNAHQALCSGNPGDMLSSNINFPINGPKR-----OMGMPMG----- 439
DB 367 lmannsifds-----gky-avsekenflftklkeepelaqlaptpgdalisl 414
OY 440 RFGSGGNHNSGMQATTPOGSNVALKMNPS--QSSPGMNGOPTSMLSPRHRMSPGVAG 498
DB 415 dfgnqfneassaygkailpqpqwatelrshstgseagslpafv-----pqaaa 464
OY 499 SPRIPPOFPSPAGSLHSPVGVCSSTGNSHSTYNSLNLQALQEGHVSGLSSLASPLK 558
DB 465 pgrtstpsatssss-----cstpnspedytstidnlik-----levleklfamdte 510
OY 559 MGNLONSFPVNNPPRLSKMGLSDSKDCFGLYGEPSEGTGQAESSHCGEOKETNDPILP 618
DB 511 akdgcstqtdfnejdlelapy-----ipmdgedfqlspc-peerllaenpqt 559
OY 619 PAVSSERADGSGRLHDSKGR-----KLQLLTTSDDMEPPLASS 660
DB 560 pq-----hcfamtnlflpavphspfilldkfqqleskktepshrmssi 607
OY 661 LSDTNKDGSTGLP--GSGSTHGTSLKRNKILNRLDSSSPVDLAKLTGAATGKDLQ 717
DB 668 ffdag--skaslppccgagastplsmggrsn-----tqwpdpplrhfpktkway-dqrt 659
OY 718 ESSSTAPGSEVITIKOEPVSPKKENALLRYLLDKDTRDGL--PEI-TPKLERLDSKTD 774
DB 660 eflgaap-----lpprvpphsvtfkr-----sakfgagpvdvlspmal----- 702
OY 775 PASNTKILAMTEKREMSFEP--ODRGSELNLF-ELIDDLONSQPLPFPDRPCAP 830
DB 703 --snkiklkrqleyeeqfqlsgdppgastshlmwkrmklrlygscp-lmpd----- 753
OY 831 AGVDKQAIINDLMQLTAENSPVPVGAQKTALRISQTFNNPRPGOLRLPNONLPLD 890
DB 754 -----xpisanvndkftg-----nmpug-1gh-plthldlp 783
OY 891 ITLOSPTGAGP-----FPPIRNSSPYSVLPQGMNGOQMGIGNOGINSSTGMIGNS 943
DB 784 ---qpasaispgenskrifppcyatlyg-----dyslssakhsqgm 822
OY 944 ASRPTMPGGEVNAPOSSAVRVCATTSAMNRPVOC 978
DB 823 asrlipgsfesylipeltlydc-----evnvpvlg 852

```

RESULT 27

AA79161

AA79161 standard; Protein; 870 AA.

AA79161;

05-JUN-2000 (first entry)

Human endothelial PAS domain protein-1.

Endothelial PAS domain protein-1; human; EPAS1; angiogenesis;

antiartherosclerotic; antitumor; atherosclerosis; tumour;  
vascular disease; vulnerability; cardiant; vasotropic;  
cerebroprotective; gene therapy; transcription factor.

Homo sapiens.

Key Location/Qualifiers  
FH 14..67  
FT Domain /note= "N-terminal bHLH domain"  
FT 486..690  
FT Domain /note= "transactivating domain"

WO200009657-A2.

24-FEB-2000.

13-AUG-1999; 99WO-US18539.

14-AUG-1998; 98US-0096515.

(HARD) HARVARD COLLEGE.

Lee M, Maemura K, Hiesh C;  
WPI; 2000-205996/18.  
N-PSDB; AA294053.

Modulation of angiogenesis in mammals, useful for treating e.g. atherosclerosis, tumors, wounds, vascular disease, hypoxic tissue damage, ischemia, balloon angioplasty, frostbite, gangrene or poor circulation

Claim 11; Page 13-14; 57pp; English.

The present sequence is that of human endothelial PAS domain protein-1 (EPAS1), a member of the transcription factor family characterized by a basic helix-loop-helix (bHLH) domain and a Per-AR-Arnt-Sim (PAS) domain. The invention is based on the discovery that EPAS1 binds to cis-acting regulatory sequences associated with genes encoding angiogenic factors such as vascular endothelial cell growth factor (VEGF) and VEGF receptors such as KDR/Flk-1 and Flt-1, thereby transactivating the promoters of such genes. A claimed method of inhibiting angiogenesis in a mammal comprises administering to the mammal a compound which inhibits binding of EPAS1 to the cis-acting transcription regulatory DNA of an angiogenic factor (see AA294051). The compound may be an EPAS1 polypeptide lacking a transactivation domain (see AA79160) or a nucleic acid encoding such a polypeptide, an antisense nucleic acid complementary to mRNA encoding EPAS1, or an EPAS1-specific antibody. The compound is preferably administered to an atherosclerotic lesion or to a tumor site. Angiogenesis is also inhibited using a compound, such as an EPAS1 dominant negative mutant, which inhibits binding of EPAS1 to the EPAS-binding element, ARNT4 (see AA79162). Angiogenesis can be promoted by administering EPAS1 DNA to increase expression of VEGF or VEGF receptor in endothelial cells of a patient suffering from peripheral vascular disease, cerebral vascular disease, hypoxic tissue damage (e.g. hypoxic damage to heart tissue), or coronary vascular disease as well as to treat patients who have, or have had, transient ischemic attacks, vascular graft surgery, balloon angioplasty, frostbite, gangrene, or poor circulation.

Sequence 870 AA;

Query Match 3.8%; Score 292; DB 21; Length 870;

Best Local Similarity 20.6%; Pred. No. 3.2e-10;

Matches 217; Conservative 138; Mismatches 406; Indels 294; Gaps 46;

10 DPSRAETRRKRECDPLSPFRNTEKRNROENKYIEELA-ELIFANFNDINDNFNPKPD 68

6 ekrsrserrke-----ksrdaarcrsketevfelhph-----svshld 52

```

OY 69 KCAILKEVYKQIROIKEQEKAAANIDEVOKSDVSTGCGIDKDALGPMMLBALDGFEE 128
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 53 kasimleisflrtkhllsvcseneseead-----qgmndlylkallegfia 100
OY 129 VVNLBGNVAVSVBNVOYLRYNOELMNKSVYLIVGHDFEVKMLLPISYNGSGWSG 188
   ||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 101 vvvqgdgmflfiseniskfmglvtqveltnhsifdfcncdeleltnl--slkngsqifg- 156
OY 189 EPPRRSHTEFNCMLVPLPDSSEEGHDNOEAHQKYEYTMQCFVNSQPKSIKEEG----- 242
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 157 -----kkskdmsterdfmfmktctntnrgtrvnlksatkvyl 193
OY 243 -----EDLOSCLICVARRVPMKERYLP--SSSEFTTRDLOLQK 279
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 194 hctgvykvynncpnhslcgykepllscillimceqlphsmdiprldsktflsrhsmdmk 253
OY 280 ITSLEDTSTRAAKPGM--EDLYRCLQKFNHAGEGESVYARHHHEVLRQGLASQIT 337
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 254 ftycdp---rteltyghpbeellgrsaayelyhaIdseentc--kslnqlctkqvvvsgy 307
OY 338 RFSLSDTVAQAOTKSKLIRSOYTNNEPOLYISLH--MLHREQNVCMNPDLTGOTWGP- 394
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 308 rmlaknggyvwtetgtvlynpnlqpqclmcyvnylsetekndvfvfsmqtr-eslflkph 366
OY 395 ---LNPISNSPAHQALCSGNPGQDMTLSSNINFPINGPRE-----QMGMPMG----- 439
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 367 lmmamslfads-----gkg-avseksnlfkklkepeelaqlaptgddalisl 414
OY 440 RFGSGGMNHSVGMQATPTPGGSNYALKMNSP--QSSPGMNPQPTSMLSRHHMSPGVAG 498
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 415 dfngnqfseesaygkallpssgpwalelrshscqseagslpaftv-----pqaa 464
OY 499 SPRIPOSPAGSLSPVGVCSSTGNSHSTNSLNAQLSEHGVSLSGSSLASPDLK 558
   ||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 465 psgtspatssss-----cstpnspedytclndlk-----ievleklfmdtce 510
OY 559 MGLQNSPVNMNPPRLSKMSGLSDKCFGLGPEPSEGTGQAESSCHPGQAKTNDPNLP 618
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 511 akqcsctqdfneldletlapy-----lpmdgedfqlsplic-peorlllaenpqst 559
OY 619 PAVSSRRADQGRRLHDSKQGT-----KLQLLTLTKSDQMEPSPPLASS 660
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 560 pq-----hcfamtnlfgplavaphspflldkfqqleskktepehrrpmssl 607
OY 661 LSDFNKDSGTSLP--GSGSTHGTSLKEKHKLHRLLDSSSPVDLAKLTAEATGKDLQ 717
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 608 ffdag--skaslppccgqastpmsmggrsn-----tqwppdpplhfgpkwvq-dqtr 659
OY 718 ESSSTPAGSEVITKQEPVSPKKENALLRYLLDKDQTKDIGL--PEI-TPKLERLDSKTD 774
   ||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 660 eflgaap-----lppvpapbhvstfktl--sakfgfargpavlspamval----- 702
OY 775 PASNTKLIAWKTEKEMSFEP--GPOPGSELBNLE-ELIDDLQNSQLPQLPFDTRPGAV 830
   ||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 703 --snklklkrqleyeeqafqdlsgdpppgstshlmwktmknltrgscp-lmpd----- 753
OY 831 AGSVNDKQAIINDMLQTLAENSPTVGAQKTALRISOSTFNNRPPQOLGRLPLNQMLPLD 890
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 754 -----kplsanvndkftcy-----nprmg-1gh-plrlhlpplp 763
OY 891 IRLQSTGAGP-----FPPINSSPYSIVITPQGMGMNGMIGNONGNLSGTGMIGNS 943
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 784 ---qpssalspgensksrtrppqcyatcyq-----dyslssahkvsvsm 822
OY 944 ASRPPTSGEMAPQSSAVRYTCAATTSAMNRPVQ 978
   ||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 823 aarl1lgsptiesyl1pel1tydc-----evnvpv1g 852

```

RESULT 28  
 ABB62819  
 ID ABB62819 standard: Protein; 1963 AA.  
 XX

```

AC ABB62819;
XX
XX 26-MAR-2002 (first entry)
DT
XX Drosophila melanogaster polypeptide SEQ ID NO 15249.
DE
XX Drosophila melanogaster polypeptide SEQ ID NO 15249.
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX Drosophila melanogaster.
XX
XX WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US09231.
XX
XX 23-MAR-2000; 2000US-191637P.
XX 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE ) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI; 2001-656860/75.
XX
XX N-PSDB; ABL06922.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX PT genes from Drosophila and for elucidating cell signalling and cell-cell
XX PT interactions -
XX
XX Disclosure; SEQ ID NO 15249; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX CC capable of detecting 1000 or more genes from Drosophila. The invention is
XX CC useful in developmental biology and in elucidating cell signalling and
XX CC cell-cell interactions in higher eukaryotes for the development of
XX CC insecticides, therapeutics and pharmaceutical drugs. The invention
XX CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
XX CC sequences (AB101840-AB16175) and the encoded proteins
XX CC (AB57737-AB72072).
XX CC The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 1963 AA;
SQ

```

Query Match 3.8%; Score 289.5; DB 22; Length 1963;  
 Best Local Similarity 20.8%; Pred. No. 1.7e-09;  
 Matches 265; Conservative 146; Mismatches 510; Indels 353; Gaps 55;

```

OY 399 SSNSPQHQAALCSGNPGQDMTLSSNINFPINGPREQGMMPGRRGSGGMNHNHSGMQATTP 458
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 155 nmnrprrtrqsalrrnrvltrannsnntppkkedlrep-----qatatla 199
OY 459 OGSNVAFLKMNPSQSSPGMNPQPTSM-----SPRRHSPGVAASPRIPSPQF 507
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 200 taataas-etasarpsavkskensltdeddaecdsdsllchkres-----psrm 250
OY 508 SPAGSLHSFPVGVCSSTGNSHSTNSLMLQAALSEHGVSLSGSSLASPDLK--MGNLON 564
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 251 rtrnkgnnnstsgnltagnngnatslssgttgggaaguns-sskqdsanavangkr 309
OY 565 SPVNMNPPRLSKKGLSDSKDCTFLYEPSEGTGQAESSCH-----PEBQKTNP 615
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 310 pkrgetlpdvsgavds-----pktrpkavessanrrkqgkqdtppkkkrtege 360
OY 616 NLPPAVSSERADQGRRLH-DSKGQTKLLOLLTTKSDQMPSPPLASSLSDFNKDSGTSLPG 674
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 361 snepsaheenalkekrrkrdspve-----smnsdrpsvlddgesnttttttaeqgs 413
OY 675 SGSTHGT-SLKEKHKLHRLLDSSSPVDLAKLTAEATGKD-----LSQSSSTAGP 725

```

```

Db 414 tksketvsckeerem-----vndleakeeekalkaalaedskdsalkmdeethiqgs 470
QY 726 SEVITKOE-----PVSPKKKENALLRLYLDKDDTKDGLPEITPKLERLDSK 772
Db 471 sadtslyvdgpnnaipspvaapilmkvptlatvealnasvdk-----ealekescs- 524
QY 773 TDPASNTKLIAKTEKEEMSFEPGDQPSSELDNLEIIDDQNSOLPOLPFDTRGAPAG 832
Db 525 -dpemlkklatikge-----vspqgqgmqqsqgqmqgqlapgyidpq--ppscp--pse 574
QY 833 SV-DKQAIINDLMQLTENSFVTVGAO-----KTALRISOSTFNPPRGOL--GRLLP 883
Db 575 svyikkepmedsmatcqnsgnepqdlkvkileknedaliksagqlppsgcappsaalp 634
QY 884 NQNLPLD-----ITLQSPFG-----AGFPPIRNSPSYSVIRPQGM 920
Db 635 lsgapvesgqglhqlqmhphqvltcqpplldqqlkypg-----sqqgvpqpqq 687
QY 921 GNOGMIGNOGNIGNSSTGMIGNSASRPTMPG-----ENAPQ-----S 958
Db 688 hs-----daagvsqagppgaptcpqkyppememktaqgdllkyppppldalrys 736
QY 959 SAVRVTCATTSAMNRPVQGMIR---NPAASIPMRPSSQPGQRQTQSOVMNIGPSEL 1014
Db 737 qemqaaaaaaakkydmkyymegqgkyvelisaahqppskpygdsikpdkpfgfhl 796
QY 1015 EMNNGGF-----QYSQO-----QAPPNQTP-----WPE 1038
Db 797 phnvgspldaahkypppptsqsgsqgqgqpqphahvppgafppglaampkphqhdvqfpp 856
QY 1039 SILPIDDASFASQOROPFGSS---PDLILCPHP---AAESPDEGALLDQLVLRNFGGL 1093
Db 857 lgrpfecqgmlkygdplaaakypgpdqkympvsgagpd-----vxy99e 905
QY 1094 EEDIRA-LGIP--ELVSQSOAVDEPQSSODSNTMLLEQKAFVFOQYAS-----QAOMA 1144
Db 906 nllkspygpppspidasarsltpqsgsqgsnsgppsmppqgqfsgpshpmpspa 965
QY 1145 QGSSTPMQDPRFHTMGORPSTATLRMQPRCLRTGLVQONPNOL-RQLDQHRLO-AQON 1202
Db 966 gggldppmbhpnllhqpypg--aaqgsqgpppppptsllhqpptsagppslqhlhpgfhl 1024
QY 1203 RQPLMNOISNVSNNVNLFLRCPVPTQA-----VOQRTLMRGGGLMNTP 1229
Db 1025 sq--lsvaasslppssiqlptlslcmaphmhlhphahlgqlhtrphldlppsmhphamp 1082
QY 1230 IMAQMLAQORREILNQHLLRQRMHQOOO-----VOQRTLMRGGGLMNTP 1274
Db 1083 lslqghpqbhqlprrshltsqgqgqgqgqgqgqgqvtrtpspaqgpprrshhdpqsrp 1142
QY 1275 SMVAPSGMPATMSNPRIPQAAAOQFPFPNNGISQDPDGTGATTPQSPLMS-----P 1328
Db 1143 pteqpsltimgssqgppp--pqgsp---nahltspilpqlagsgpppqlilghpmaalp 1196
QY 1329 RMAHTQSPMMQSOANPAYOA-----PSDINGW-----AOCNMG--NSM 1366
Db 1197 hlalhl-----ppghpbaaalhphghllshslaglgpggqplalagpagaigalpesa 1249
QY 1367 FSQOSPP-----HFQOQANT-----SMYSNNMNINVMATNTGMSMNOMTGOISMTS 1415
Db 1250 lstrtppshphshasapllahvasamtslsm-----slstsvpsafstraspsvqlss 1305
QY 1416 VTSVSTSGLSMGP 1429
Db 1306 -sgqgsppsgpsvvp 1318

```

RESULT 29

AAW68094

ID AAW68094 standard; Protein; 816 AA.

XX AAW68094;

AC

```

XX 09-NOV-1998 (first entry)
DT Mouse neuronal PAS domain protein NPAS2.
XX NPAS2; neuronal bHLH-PAS domain; mouse; transcription factor;
XX therapy; diagnosis.
XX Mus sp.
XX W09831804-A1.
XX 23-JUL-1998.
XX 21-JAN-1998; 98WO-US01154.
XX 21-JAN-1997; 97US-0785310.
XX (TEXA ) UNIV TEXAS SYSTEM.
XX McKnight SL, Russel DW;
XX WPI; 1998-414103/35.
XX N-PSDB; AAV41259.
XX New isolated neuronal PAS domain proteins - can regulate function of
XX neurological tissue such as brain tissue, used to develop products
XX for diagnosis and therapy
XX Claim 1; Page 24-37; 42pp; English.
XX
XX This is the amino acid sequence of mouse neuronal PAS domain protein
XX NPAS2, a new member of the basic helix-loop-helix (bHLH)-PAS family
XX of transcription factors. A database search for expressed sequence
XX tags bearing sequence similarity to the PAS domain of the aryl
XX hydrocarbon receptor yielded 2 clones. These were used to generate
XX primers for PCR amplification of hybridisation probes, and clones
XX (see AAV41245-59) coding for human and mouse NPAS proteins (see
XX AAW68091-94) were isolated from mouse brain tissue, human brain
XX tissue and HeLa cells. The NPAS proteins can regulate the function
XX of neurological tissue such as brain tissue. They can be produced
XX recombinantly from transformed host cells or purified from
XX mammalian cells. NPAS proteins and polynucleotides can be used in
XX diagnosis (e.g. genetic hybridisation screens for NPAS transcripts),
XX therapy (e.g. gene therapy to modulate NPAS gene expression) and in
XX the biopharmaceutical industry (e.g. as immunogens, reagents for
XX isolating other transcriptional regulators, and reagents for
XX screening chemical libraries for lead pharmacological agents).
XX
XX Sequence 816 AA;
XX
XX Query Match 3.8%; Score 289; DB 19; Length 816;
XX Best Local Similarity 18.3%; Pred. No. 4.6e-10;
XX Matches 211; Conservative 151; Mismatches 360; Indels 432; Gaps 42;
XX
XX 4 MGENTSPDSRAETKRKRCQPOLQPSRKNTKRNREDEKNTIELAELIRANNDIDNF 63
XX 1 mdedekdrakrasrnk-----sekrrtdqfnvllkelsmslpgntr----- 41
XX
XX 64 NFKPDKCAILKEVTKOIRKEOEKAAANIDEVOKSOVSSTGGGVIDKDLGPMLEAL 123
XX 42 --kmdktvlekvigflg--khnevsagtelcidqgdtkps-----flsneftqlmeal 93
XX
XX 124 DGEFFVNLGNNVVFSENVTOYLRYNOELMNSVSYSLHVGDTHEFVKNLPRKSIYNG 183
XX 94 dgflvltcdgsllyvsdstlpllglnhpadvmdqnlInflpeghsevykllsshmIvtd 153
XX
XX 184 GSNSEPPRRNSHTFNCML---VKP-----LPDSEEBGHDN--- 217
XX
XX 154 spspeflksdndlefyhchllirgslnpkfeyikfyvgnfrsynnpsscngfdntls 213
XX
XX 218 OEAHQKYTEMQCF-----AVSQPKSIKEGEDLIQCLICVARRVPMKRPVLPSSSEFTTR 273

```



```

Db      214 rpehprlgkvcfistatvratpqlfke-----mcva-----depl-----eeftsr 254
OY      274 ODLOCKITSLDSTWRAAKPGM---EDLVRRCIQKFNHQHGESVYAKRHHHEVLRQG 330
Db      255 hslwffild---hrappiigy1pfevlgtsygnynhl-----dallarchqhmfig 307
OY      331 LAFSQTIRFSLSDGTLVAAQTSKILRSQTNPEQLVLSH-----MLHREQNVCM 382
Db      308 kqkscgyrlftcggqwtwqthnylytyhgmakpelfcthvsvadydrvertrqelale 367
OY      383 NPDLTGQTMGKPLNPISNSPAHQALCSGNPGQDMTLSSNINFPINGPEQGMWGRFG 442
Db      368 qp-----pleam----- 374
OY      443 GSGGMNHVSGMQATTPOGNSYALKMNSPQSGMNGOPTSMLSPRHMSGVAGSPRI 502
Db      375 -----hpsavk-----ekdsleppqpfnal-----dmgaaglpss 405
OY      503 PPSQFSPAGSLSPVGVCSSTGNSHSYTNSSINALQALSEGHGVSLGSSLASPDLMKGNL 562
Db      406 p-----spasars-----shkshtlams----- 424
OY      563 QNSPVNMNPPPLSKMGLSDSKDQFGLYGER-SBGTGQAESSCHEQEKEETNDPNIIPAV 621
Db      425 -----eptstptklmaenstlaipratlqpqlpvag 456
OY      622 SSERADGOSRLDSKG---QTKLQLLTTSKSDQMEPSPLAS-----SLSDNKGDTGSL 672
Db      457 lsgaaampalhsaaadltkqlllqslpqtqglsppavvtqfsagfmftqtkd----- 511
OY      673 PGSGSTHGTSLKEKHILHRLLODSSSPVDLAKLTAEATGKDLSDSSSTAPGSEVTIKO 732
Db      512 -----qlqgrtrllq-----anlrwgq 528
OY      733 EYVSPKKKRNALLRYLLBDMDTKDGLPEITPKLERLDSKTDPASTKLIAMKTEKEEMS 792
Db      529 eelhkiqueqlclvq-----dsnvqmf-lqqpavsls 558
OY      793 FEPGDQPGSELNLEETLDDLIQNSQLPOLFPTRPCAPAGSVDKQAITINDLMQLTAEENS 852
Db      559 fssqtqpaaq-----qqllqg-----rpaaps-----qpqlvvn----- 586
OY      853 VTPVGAQKTALRISQS-----TFNNPRPGQLGRLLPNQNLPL-DITLQSPTGAGPF 902
Db      587 -lplgqgltstqvtqnhllresnvlsagqpkpmtsrsgllpasgrslslpsqfscstasvl 645
OY      903 PPIRNSPPTSVIPQPGMNGOGMIGNGLGNS-STGMIGNSASRPTMSEGAAPQSSAV 961
Db      646 pqlsliltiaptq-----dsqcgpspdtghdrqlrlllspqlqpmmpqscdarpsav 700
OY      962 RVT-----CAATTSAMNRPVQGMIRMPASIPRPSQPGQROTLQS 1004
Db      701 strtgrykvkagqymfmpsdshptnsastpv---llmgqavlhpsfpastrpslq----- 753
OY      1005 QVWNIGPSELNMNGPQYSGQOAPNQTAP-----WPESILPIDQASFASQNRQPFSS 1058
Db      754 -----paqagqgpppylqaptalhsqpsdl-----lftsqppqtlqy 793
OY      1059 SPDDLCPHPAAS 1072
Db      794 aatqslcpqpprps 807

```

```

XX      KW Drosophila: developmental biology; cell signalling; insecticide;
XX      KW pharmaceutical.
XX      OS Drosophila melanogaster.
XX      PN WO200171042-A2.
XX      PD 27-SEP-2001.
XX      PF 23-MAR-2001; 2001WO-US09231.
XX      PR 23-MAR-2000; 2000US-191637P.
XX      PR 11-JUL-2000; 2000US-0614150.
XX      PA (PEKE ) PE CORP NY.
XX      PI Venter JC, Adams M, Li PWD, Myers EW;
XX      DR WPI, 2001-656860/75.
XX      DR N-PSDB; ABL06334.
XX      PT New isolated nucleic acid detection reagent for detecting 1000 or more
XX      PT genes from Drosophila and for elucidating cell signalling and cell-cell
XX      PT interactions -
XX      PS Disclosure: SEQ ID NO 13485; 21pp + Sequence Listing; English.
XX      CC The invention relates to an isolated nucleic acid detection reagent
XX      CC capable of detecting 1000 or more genes from Drosophila. The invention is
XX      CC useful in developmental biology and in elucidating cell signalling and
XX      CC cell-cell interactions in higher eukaryotes for the development of
XX      CC insecticides, therapeutics and pharmaceutical drugs. The invention
XX      CC discloses genomic DNA sequences (ABLI6176-ABLI30511), expressed DNA
XX      CC sequences (ABLI01840-ABLI6175) and the encoded proteins
XX      CC (ABBS7737-ABB72072).
XX      CC The sequence data for this patent did not form part of the printed
XX      CC specification, but was obtained in electronic format directly from WIPO
XX      CC at ftp.wipo.int/pub/published_pcl_sequences.
XX      CC Sequence 2441 AA;
XX      SQ

```

Query Match 3.8%; Score 287.5; DB 22; Length 2441;  
Best Local Similarity 19.2%; Pred. No. 3.2e-09;  
Matches 296; Conservative 195; Mismatches 595; Indels 453; Gaps 61;

```

OY      6 EMTSDPSR-----AETPKRKECPDQLGSPKRTKRNREOEKYEIELLAFANPDID 61
Db      1136 ertresrkaaaarrrerk-----kkmekkeekrrqgngpgddmqgdddasdkd 1190
OY      62 NFNFKPDKCALIKETVKQIRQIKEDKKAANIDEVQKSDVSTGCGVTDKDAL--GPM 119
Db      1191 ddsdkd-----edeaapaaree-----gdsjldgscssgdkd 1226
OY      120 LEALDGEFFVNLVGNVVFSENVTOYLRYNOBELMKNKSVYSLIHVGHTFEVKNILPKS 179
Db      1227 garfgsgsaagaanaansvtn--sgkknkqaknkylisv---eptqvy--ltsns 1278
OY      180 IYNGSGMSGEPP-----RMSHTFNCMLVYKPLPDSSEEGHDNOEAHQKETNQCFVNSQP 235
Db      1279 vlkgycaakhhpavevkkqpapatqaaqlkrqldvkkee-----palkkkekns---sss 1330
OY      236 KSIKEGEDLQSLICVARVMPKKEP-----VLPSSEFTTRQDLOGKITSLDTSTMRAL 291
Db      1331 ssakrekenlapkava-----lpkqgsssssklgsesasinlstatltsaantlrkey 1386
OY      292 MKP-----GMEDLVRRCIQK-----FHAQHGESVSYA 319
Db      1387 akpasqlasatltnpakrtrevdgwkvevrkksaqtlaygaagaprlpvatasatsvqnh 1446
OY      320 KRHHHEVLRQGLAFSQTIRFSLSDGTLVAAQTSKILRSQTNPEQLVLSLHMLHREQNV 379

```

```

Db 1447 phhh-----lans-----ssnssslttsttaassv----- 1473
Qy 380 CVMNPDLTGTMGRPLNFISNSPAHQALCSGNPGDWTLSNINFINPGCKEQGMGMDG 439
Db 1474 -----pemtckkvpyvnaais----- 1489
Qy 440 RFGSGGNNHVSQAOATTPOGSSNVALKMNPSOSSPGMNPQOPTSMLSPRRHMSPGVAGS 499
Db 1490 rviqrgsgn-inairatl-gahlevekqkngserctllkyltdatqahmlialikd 1546
Qy 500 PRIPPSQFSPAGSLSPVGVCSSTGNSHSYTNSSLNALQALSEHGVSLSASLPDLKM 559
Db 1547 pdvdlqmlpr-----insstik-qasssg-----astpmav 1576
Qy 560 GNLGN-SPVNNNPPLSKMGLSKDCPGLXGPESEGT--GOAESCHPEOKETNP 616
Db 1577 gtwmrtcaagynaytfssaaatstss-----sssaasttpgaasyshahknhqdgpsvk 1632
Qy 617 LPPAVSE--RADQS--RLHDSKQTKLLQLLTTRKSDMPEPSPLASSLSDTNKDSGSLP 673
Db 1633 gpgsgstsvkngsstkvaasags-----gsrsgagssyla-----gqpgjrss 1678
Qy 674 GSGSTHGTSLKEKKHLRLHLDSSSPVDLAKLAETAGKLSOBSSSTAGSEVTTKOE 733
Db 1679 gggssngv-lksks-----essskslpaakksstlqk-----sstlvspgaqnfaakaa 1725
Qy 734 PV---SPKKKENALLRYLLDKDTRKDGLEPITPKLERLDSTKTPA----- 776
Db 1726 aigsgspkkaeggaatsavvtaagrsqgvapfgrgkpvaggggaataaasnavqlgsvs 1785
Qy 777 -----SNTKLIA----- 783
Db 1786 gsnllagpigtfnvadaavnaaagaataasnvkrlapiapskrvgsptqvqgh 1845
Qy 784 -MKTEKEMSEFEPDQSGSELDNLEHLLDQLNSQLPOLFPD--TRGAPAGS----YDK 836
Db 1846 qtcqqqqqqlpqpapvypqpq--qbp1qgqqqqqabqpqqpqqpqtsgnqlvin 1903
Qy 837 QAILNDLMOLTAENSPTVPVGAOKTA--LRISQSTFNNRPOLGRLLPQNQLPDLITLOS 895
Db 1904 tnllndlmaasaanttsdtsfaqklaaklsayslfsdyqsgqgkl----- 1949
Qy 896 PTGAGPPPTLRNSPYVIOPGMGNOGMIGN--QCNLGNSTGMIIGNSASRPTMFG 952
Db 1950 -----qdpq1gqgagavdg1pqad-askapgyurnllspsvgsk 1989
Qy 953 EMAPQSSAVRYTCATISAMNRPVQGMIRNPASIPMRBSQCGQKOTLOSQVNNIGPS 1012
Db 1990 assnhsctspvgynv1qgqqqqqpqsqqa1nlltspgypataparapmwsaanqn--pa 2047
Qy 1013 ELENNMGCPQYSQOQAP---PNTGAPWPESTILP1DOASFSASQNPQPGSSPDDLCP--- 1066
Db 2048 vqgqsmgtqglgetapahspgv1kp-pta1vpl-----qnhvmp1sepegapptf 2099
Qy 1067 -----HA-----AESPSDEGALLDLYLARNFDELEIDRALG1PELVSOSO-- 1110
Db 2100 gajsgnpsagnsaagaataaasamltdcqqnqlqltqlnqlmvgasqgqqpqq1 2159
Qy 1111 --ANDP--EQFSSQDSN1MLQKAVFPQOYASQAOAMOGS1PMQDNFRHMGCRPSYAT 1167
Db 2160 nypmdfcsst1vdannv1rl1nprv1fp-----qgn1ckppqp-----pqg1t 2201
Qy 1168 LR--MQRPEGLRPTGLVQNPQNLRLQ-----LOHRLQAOQNPQPLMNOISVSNVLT 1219
Db 2202 qsnvfgnpprqpptg1arqpggaagqrwyg1tleypsygrdm1lhengaggaagm--- 2258
Qy 1220 LRGEVPTQAD1NAOMLAQORREILNQHLRQRQMHQOQOVQORTLMMGQGLNMTPSNAP 1279
Db 2259 ---gspsamspnhd1rkmprijt1e-----raaswkyntfnvgps1nmedala-- 2305
Qy 1280 SGMFATNSNRIP1QANNAOQPFPPNVTISQOPDGF1GATTP1QSLSPRAH1QSPM1Q 1339
Db 2306 sv1ppwhe1k1kaqpg1qdp1pppsqgqqqplnw-----1kqbp1qgqy1raynngpyq 2361

```

```

Qy 1340 QSOANPAYQAPSDINGMAOGNMGNSMFSQSPPHFGQO 1378
Db 2362 qgqgqhepmmpmdyh-----nmqapmmsq 2387

```

```

RESULT 31
ABB31417
ID ABB31417 standard; Peptide; 149 AA.
AC ABB31417;
DT 01-FEB-2002 (first entry)
DE Peptide #4068 encoded by breast cell single exon nucleic acid probe.
XX Human; microarray; single exon probe; gene expression; breast;
KW disease; cancer.
XX Homo sapiens.
XX W0200157271-A2.
XX 09-AUG-2001.
XX 30-JAN-2001; 2001WO-US00662.
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-496933/54.
XX New spatially-addressable set of single exon nucleic acid probes,
XX useful for measuring gene expression in sample derived from human
XX breast, comprises number of single exon nucleic acid probes -
XX Claim 27; SEQ ID NO 14385; 327pp + sequence listing; English.
XX
XX The invention relates to a spatially-addressable set of single exon
XX nucleic acid probes for measuring gene expression in a sample derived
XX from human breast and BR 474 cells. The method involves contacting
XX the probes with a collection of detectably labelled nucleic acids
XX derived from mRNA of human breast, and then measuring the label
XX bound to each probe of the microarray. The probes are useful for
XX verifying the expression of regions of genomic DNA predicted to
XX encode proteins. They are useful for gene discovery, and for
XX determining predisposition and/or prognosing breast disease. Gene
XX expression analysis is useful for assessing the toxicity of chemical
XX agents on cells. The microarray of this invention presents a far greater
XX diversity of probes for measuring gene expression, with far less bias
XX than expressed sequence tag microarrays. The method is suitable for
XX rapid production of functional information from genomic sequence. The
XX present sequence is a peptide encoded by a single exon nucleic acid
XX probe of the invention.
XX Note: The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pcl_sequences.
XX
XX Sequence 149 AA:

```

```

Query Match 3.88; Score 287; DB 22; Length 149;
Best Local Similarity 44.28; Pred. No. 4.4e-11;
Matches 61; Conservative 31; Mismatches 42; Indels 4; Gaps 2;

```

Oy		653	EPSPLASSLSDTNMDSNCSLGGSSSTGCTSLKEKHKLHLNLDDSSSPVLAFLTAATG	712
Oy			: : : : :     :     :     :     :	
Dd		5	ESSVSVTPSPGVSSSTSGVVSSTSMHGSLLQEKHRLHLKLLQGNSPAEVAKILAEATY	64
Oy		713	KDLSSSESTFAPGSEVTRTKQPVSPKKKE-NALLRVLTDKDPTDGLPETTPKLERLDS	771
Oy			: : : : :     :     :     :     :	
Dd		65	kD---tssitcgcgdnvvkqeqilppkkennallrlylldrddpsdaIskeIqpqvgevh	121
Oy		772	KTDPASNTKIAMKTEKE	789
Oy			: : : : :	
Dd		122	kmsgtstslpssseqkd	139
RESULT		32		
ABB36630		ID	ABB36630 standard; Peptide: 149 AA.	
XX		AC	ABB36630;	
XX		DT	04-FEB-2002 (first entry)	
DE		Xx	Peptide #4136 encoded by human foetal liver single exon probe.	
XX		KW	Human; foetal liver; gene expression; single exon nucleic acid probe.	
OS		Xx	Homo sapiens.	
PN		XX	WO200157277-A2.	
PD		XX	09-AUG-2001.	
PE		XX	30-JAN-2001; 2001WO-USO0669.	
PR		XX	04-FEB-2000; 2000US-0180312.	
PR		XX	26-MAY-2000; 2000US-0207456.	
PR		XX	30-JUN-2000; 2000US-0608408.	
PR		XX	03-AUG-2000; 2000US-0632366.	
PR		XX	21-SEP-2000; 2000US-0234687.	
PR		XX	27-SEP-2000; 2000US-0236359.	
XX		XX	04-OCT-2000; 2000GB-0024263.	
PA		(MOLE-)	MOLECULAR DYNAMICS INC.	
PI		Penn SG,	Hanzel DK, Chen W, Rank DR;	
XX		XX	WPI; 2001-483447/52.	
PT		XX	Human genome-derived single exon nucleic acid probes useful for	
PT		XX	analyzing gene expression in human fetal liver -	
PS		XX	Claim 27; SEQ ID NO 29265; 639pp + sequence listing; English.	
CC		XX	The invention relates to a single exon nucleic acid probe for	
CC		XX	measuring human gene expression in a sample derived from human foetal	
CC		XX	liver. The single exon nucleic acid probes may be used for predicting	
CC		XX	measuring and displaying gene expression in samples derived from human	
CC		XX	fetal liver. The present sequence is a peptide encoded by a single exo	
CC		XX	nucleic acid probe of the invention.	
CC		Note:	The sequence data for this patent did not form part of the	
CC		printed specification, but was obtained in electronic format directly		
CC		from WIPO at ftp.wipo.int/pub/published_pct_sequences.		
Sequence		149 AA;		

Query Match	3.8%	Score 267;	DB 22;	Length 149;
Similarity	44.2%	Pred. No.	4.4e-11;	
Best Local	Conservative	31;	Mismatches 42;	Indels 4; Gaps 2
Oy	653	EEPSPLASSLSDPNKSTGSLPGSGSTHGTSLKEKHRLIHLRLQDSSSPDLAKLTAEATNG	712	
db	5	esssvetvcpvgysstsgvssstsmhbgslqeknhlhlkllqnqngspveevklkeatga	64	

```
OY      713 KDLSSESSSTARPSGEVITKOEPPSPKKKE -NALLRRLYLLDDDTFDIGLPTETFKLEMLDS 771
OY      65 kd---tssticggnvvgkqqlspkckennallrlylldrddpsdalskelqpvegvdm 121
Db
OY      772 KTDPA$NTKLIA$MTEKE 789
Db      122 kmsqctsetslppssqekd 139
```

RESULT 33  
 ID ABB21964  
 XX ABB21964 standard; Protein; 149 AA.  
 XX  
 AC ABB21964;  
 XX  
 DT 23-JAN-2002 (first entry)  
 XX  
 DE Protein #3963 encoded by probe for measuring heart cell gene expression.  
 XX  
 KW Human; gene expression; heart; microarray; vascular system;  
 KW cardiovascular disease; hypertension; cardiac arrhythmia;  
 KW congenital heart disease.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200157274-A2.  
 XX  
 PD 09-AUG-2001.  
 XX  
 PF 30-JAN-2001; 2001WO-US000666.  
 XX  
 PR 04-FEB-2000; 2000US-0180312.  
 PR 26-MAY-2000; 2000US-0207456.  
 PR 30-JUN-2000; 2000US-0608408.  
 PR 03-AUG-2000; 2000US-0632366.  
 PR 21-SEP-2000; 2000US-0234687.  
 PR 27-SEP-2000; 2000US-0236359.  
 PR 04-OCT-2000; 2000GB-0024263.  
 XX  
 PA (MOLE-) MOLECULAR DYNAMICS INC.  
 XX  
 PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX  
 DR WPI; 2001-488899/53.  
 XX  
 PT Single exon nucleic acid probes for analyzing gene expression in human  
 PS hearts -  
 XX  
 PS Claim 15; SEQ ID NO 23734; 530pp; English.  
 XX  
 CC The present invention relates to single exon nucleic acid probes for  
 CC measuring human gene expression in a sample derived from human heart (see  
 CC ABA21555-ABA11305). The present sequence is a protein encoded by one such  
 CC probe. The probes may be used for predicting, measuring and displaying  
 CC gene expression in samples derived from the human heart via microarrays.  
 CC By measuring gene expression, the probes are useful for predicting,  
 CC diagnosing, grading, staging, monitoring and prognosing diseases of the  
 CC human heart and vascular system e.g. cardiovascular disease,  
 CC hypertension, cardiac arrhythmias and congenital heart disease.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 149 AA;  
 XX

Query Match	3.88;	Score 287;	DB 22;	Length 149;
Best Local Similarity	44.2%;	Pred. No. 4.4e-11;		
Matches	61;	Mismatches 42;	Indels 4;	Gaps 2
0Y	653	EPSPLASSLSDTNKDSGLPSGSGTHGTSLSKEKHRIHRLLDSSSPYDLAKLTAAATG	712	

```

Db      5  essvsvtspgsvssstsgvstsmhgsllqekhrllhkl1lqngnspeavakltaeatg 64
Oy      713 KDLQSSSTFADGSEVTKQEPVSPKKE-NALLRYLLDKDDTKDIGLPEITPKLERDS 771
Db      65 kd---tsitscggdgnvvkqglspkkennallryllldrdpsdalskelqpvegydn 121
Oy      772 KTDPAASNTKLIAMKTEKE 789
Db      122 kmsqctstlpsseqkd 139

RESULT 34
AAM57393
ID      AAM57393 standard; Protein: 149 AA.
XX
AC      AAM57393;
XX
DT      05-NOV-2001 (first entry)
XX
DE      Human brain expressed single exon probe encoded protein SEQ ID NO: 29498.
XX
KW      Human; brain expressed exon; gene expression analysis; probe;
KW      microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW      epilepsy; cancer.
XX
OS      Homo sapiens.
XX
PN      WO200157275-A2.
XX
PD      09-AUG-2001.
XX
PF      30-JAN-2001; 2001WO-US00667.
XX
PR      04-FEB-2000; 2000US-0180312.
PR      26-MAY-2000; 2000US-0207456.
PR      30-JUN-2000; 2000US-0608408.
PR      03-AUG-2000; 2000US-0632366.
PR      21-SEP-2000; 2000US-0234687.
PR      27-SEP-2000; 2000US-0236359.
PR      04-OCT-2000; 2000GB-0024263.
XX
PA      (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI      Penn SG, Hanzel DK, Chen W, Rank DR;
XX
WPI: 2001-483446/52.
XX
Single exon nucleic acid probes for analyzing gene expression in human
PT      brains -
XX
PS      Example 4; SEQ ID NO: 29498; 650pp + Sequence Listing; English.
XX
CC      The present invention provides a number of single exon nucleic acid
CC      probes which are derived from genomic sequences expressed in the human
CC      brain. They can be used to measure gene expression in brain cell samples,
CC      which may enable the diagnosis and improved treatment of nervous system
CC      diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC      epilepsy and cancers. The present sequence is a protein encoded by one of
CC      the probes of the invention.
XX
SQ      Sequence 149 AA;

Query Match      3.88; Score 287; DB 22; Length 149;
Best Local Similarity 44.28; Pred. No. 4.4e-11;
Matches 61; Conservative 31; Mismatches 42; Indels 4; Gaps 2;

Oy      653 EPSPLASSLSDTNKDSGLPGSGSTHGTSLKEKHILHRLLODSSSPVDLAKLTAATG 712
Db      5  essvsvtspgsvssstsgvstsmhgsllqekhrllhkl1lqngnspeavakltaeatg 64
Oy      713 KDLQSSSTFADGSEVTKQEPVSPKKE-NALLRYLLDKDDTKDIGLPEITPKLERDS 771
Db      122 kmsqctstlpsseqkd 139

```

```

Db      65 kd---tsitscggdgnvvkqglspkkennallryllldrdpsdalskelqpvegydn 121
Oy      772 KTDPAASNTKLIAMKTEKE 789
Db      122 kmsqctstlpsseqkd 139

RESULT 35
AAM69785
ID      AAM69785 standard; Protein: 149 AA.
XX
AC      AAM69785;
XX
DT      06-NOV-2001 (first entry)
XX
DE      Human bone marrow expressed probe encoded protein SEQ ID NO: 30091.
XX
KW      Human; bone marrow expressed exon; gene expression analysis; probe;
KW      microarray; cancer; leukaemia; lymphoma; myeloma.
XX
OS      Homo sapiens.
XX
PN      WO200157276-A2.
XX
PD      09-AUG-2001.
XX
PF      30-JAN-2001; 2001WO-US00668.
XX
PR      04-FEB-2000; 2000US-0180312.
PR      26-MAY-2000; 2000US-0207456.
PR      30-JUN-2000; 2000US-0608408.
PR      03-AUG-2000; 2000US-0632366.
PR      21-SEP-2000; 2000US-0234687.
PR      27-SEP-2000; 2000US-0236359.
PR      04-OCT-2000; 2000GB-0024263.
XX
PA      (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI      Penn SG, Hanzel DK, Chen W, Rank DR;
XX
WPI: 2001-488900/53.
XX
Human genome-derived single exon nucleic acid probes useful for
PT      analyzing gene expression in human bone marrow -
XX
PS      Example 4; SEQ ID NO: 30091; 658bp + Sequence Listing; English.
XX
CC      The present invention provides a number of single exon nucleic acid
CC      probes which are derived from genomic sequences expressed in the human
CC      bone marrow. They can be used to measure gene expression in bone marrow
CC      samples, which may enable the improved diagnosis and treatment of cancers
CC      such as lymphoma, leukemia and myeloma. The present sequence is a
CC      protein encoded by one of the probes of the invention.
XX
SQ      Sequence 149 AA;

Query Match      3.88; Score 287; DB 22; Length 149;
Best Local Similarity 44.28; Pred. No. 4.4e-11;
Matches 61; Conservative 31; Mismatches 42; Indels 4; Gaps 2;

Oy      653 EPSPLASSLSDTNKDSGLPGSGSTHGTSLKEKHILHRLLODSSSPVDLAKLTAATG 712
Db      5  essvsvtspgsvssstsgvstsmhgsllqekhrllhkl1lqngnspeavakltaeatg 64
Oy      713 KDLQSSSTFADGSEVTKQEPVSPKKE-NALLRYLLDKDDTKDIGLPEITPKLERDS 771
Db      65 kd---tsitscggdgnvvkqglspkkennallryllldrdpsdalskelqpvegydn 121
Oy      772 KTDPAASNTKLIAMKTEKE 789
Db      122 kmsqctstlpsseqkd 139

```

## RESULT 36

AAM17609

ID AAM17609 standard; Protein; 149 AA.

AC AAM17609;

DT 12-OCT-2001 (first entry)

DE Peptide #4043 encoded by probe for measuring cervical gene expression.

KW Probe; human; microarray; gene expression; cervical epithelial cell; cervical cancer.

OS Homo sapiens.

PN MO200157278-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00670.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488901/53.

XX Human genome-derived single exon nucleic acid probes useful for

XX analyzing gene expression in human cervical epithelial cells -

XX Claim 27; SEQ ID No 22435; 487bp; English.

XX The present invention relates to human single exon nucleic acid probes

XX (SENPs; see A1110068-A112459). The present sequence is a peptide encoded

XX by one such probe. The SENPs are derived from human HeLa cells. The SENPs

XX can be used to produce a single exon microarray, which can be used for

XX measuring human gene expression in a sample derived from human cervical

XX epithelial cells. By measuring gene expression, the probes are therefore

XX useful in grading and/or staging of diseases of the cervix, notably

XX cervical cancer.

XX Note: The sequence data for this patent did not form part of the printed

XX specification, but was obtained in electronic format directly from WIPO

XX at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 149 AA:

SQ

Query Match 3.8%; Score 287; DB 22; Length 149;  
Best Local Similarity 44.2%; Pred. No. 4.4e-11;  
Matches 61; Conservative 31; Mismatches 42; Indels 4; Gaps 2;

QY 653 EPSPPLASSLSDTNKDSGSLPGSGSTHGTSLKSKKHLRLHLLDSSSPVDLAKTAEATG 712

DB 5 ESSSVYTSIPSGVSSSTSGVSSSTSMHSGSLGKHLRLHLLDSSSPVDLAKTAEATG 64

QY 713 KDSLOESSSTAPGSEVITIKOEPVSPKKE-NALLRVLDDKDDPKDIDGLPEITPKLERLDS 771

DB 65 Kd---tsstiscgdgnvvkqegispkkekennallrylltdcpdsalskelqpvegvdn 121

QY 772 KTDPASTKLIAMKTEKE 789

DB 122 kmsqctsltpsssgkd 139

## RESULT 37

AAM30127

ID AAM30127 standard; Protein; 149 AA.

AC AAM30127;

DT 17-OCT-2001 (first entry)

DE Peptide #4164 encoded by probe for measuring placental gene expression.

KW Probe; microarray; human; placenta; antenatal diagnosis;

OS Homo sapiens.

PN MO200157272-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00663.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-48897/53.

XX Human genome-derived single exon nucleic acid probes useful for

XX analyzing gene expression in human placenta -

XX Claim 27; SEQ ID No 30396; 654bp; English.

XX The present invention relates to single exon nucleic acid probes (SENPs;

XX see A1131315-A1157546). The present sequence is a peptide encoded by one

XX such probe. The probes are useful for producing a microarray for

XX predicting, measuring and displaying gene expression in samples derived

XX from human placenta. The probes are useful for antenatal diagnosis of

XX human genetic disorders.

XX Sequence 149 AA:

SQ

Query Match 3.8%; Score 287; DB 22; Length 149;  
Best Local Similarity 44.2%; Pred. No. 4.4e-11;  
Matches 61; Conservative 31; Mismatches 42; Indels 4; Gaps 2;

QY 653 EPSPPLASSLSDTNKDSGSLPGSGSTHGTSLKSKKHLRLHLLDSSSPVDLAKTAEATG 712

DB 5 ESSSVYTSIPSGVSSSTSGVSSSTSMHSGSLGKHLRLHLLDSSSPVDLAKTAEATG 64

QY 713 KDSLOESSSTAPGSEVITIKOEPVSPKKE-NALLRVLDDKDDPKDIDGLPEITPKLERLDS 771

DB 65 Kd---tsstiscgdgnvvkqegispkkekennallrylltdcpdsalskelqpvegvdn 121

QY 772 KTDPASTKLIAMKTEKE 789

DB 122 kmsqctsltpsssgkd 139

## RESULT 38

AAM05268

ID AAM05268 standard; Protein; 149 AA.

AC AAM05268;

DT 09-OCT-2001 (first entry)  
XX Peptide #3950 encoded by probe for measuring breast gene expression.  
DE  
XX Probe: human; breast disease; breast cancer; development disorder;  
KW Inflammatory disease; proliferative breast disease; non-carcinoma tumour.  
XX Homo sapiens.  
OS  
PN WO200157270-A2.  
PD 09-AUG-2001.  
PF 29-JAN-2001; 2001WO-US00661.  
XX  
PR 04-FEB-2000; 2000US-0180312.  
PR 26-MAY-2000; 2000US-0207456.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2001-476286/51.  
DR  
XX Novel single exon nucleic acid probe used to measuring gene expression  
PT in a human breast -  
XX  
PS Claim 27; SEQ ID NO 14008; 322pp; English.  
XX  
CC The present invention relates to novel single exon nucleic acid probes  
CC (see AAI00010-AA110067). The present sequence is a peptide encoded by one  
CC such probe. The probes are useful for measuring human gene expression in  
CC a human breast sample, where the probe hybridises at high stringency to a  
CC nucleic acid expressed in the human breast. The probes are useful for  
CC predicting, diagnosing, grading, staging, monitoring and prognosing  
CC diseases of the human breast, particularly those diseases with polygenic  
CC aetiology. The diseases include: breast cancer, disorders of development,  
CC inflammatory diseases of the breast, fibrocystic changes, proliferative  
CC breast disease and non-carcinoma tumours.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pcl\_sequences.  
XX  
SQ Sequence 149 AA;  
XX  
Query Match 3.8%; Score 287; DB 22; Length 149;  
Best Local Similarity 44.2%; Pred. No. 4.4e-11;  
Matches 61; Conservative 31; Mismatches 42; Indels 4; Gaps 2;  
XX  
QY 653 EPSPLASSLSDTNKDSGLPSGSGSTGTSLEKHKIILRLDPSPPVDIAKLTAATG 712  
DB 5 ESSVSVTSPEVSSTSGVSTSNMHSLLGKHLRIHKLILGNSPAEVAKIAEATG 64  
QY 713 KDLSESSSTAPGSEVTIKQEPVSPKKKE-NALLRYLLDKDPTKDIGLPETPKLERADS 771  
DB 65 kd--tsstiscgqgnvkvqslpkkkennallryllrddpdsalskeiqpvegydn 121  
QY 772 KTDPASTKILAMTEKE 789  
DB 122 kmsqctslipssqekd 139

XX  
DT 07-NOV-2001 (first entry)  
XX Human novel secreted protein, Seq ID 1139.  
DE  
XX Human; immunosuppressive; antiarthritic; antirheumatic;  
KW cytosolatic; cardiant; vasotrophic; cerebroprotective; nootropic;  
KW neuroprotective; antibacterial; virucide; fungicide; ophthalmological;  
KW vulnerrary; secreted protein; rheumatoid arthritis;  
KW hyperproliferative disorder; cardiovascular disorder; cardiac arrest;  
KW cerebrovascular disorder; cerebral ischaemia; angiogenesis;  
KW nervous system disorder; Alzheimer's disease; infection; ocular disorder;  
KW corneal infection; wound healing; epithelial cell proliferation;  
KW skin ageing; food additive; preservative; antiproliferative.  
XX  
OS Homo sapiens.  
PN WO20015322-A2.  
PD 02-AUG-2001.  
PF 17-JAN-2001; 2001WO-US01341.  
XX  
PR 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198013.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226868.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.

PR	18-SEP-2000	2000US-0233081
PR	18-SEP-2000	2000US-0233196
PR	14-SEP-2000	2000US-0233397
PR	14-SEP-2000	2000US-0233398
PR	14-SEP-2000	2000US-0233399
PR	14-SEP-2000	2000US-0233400
PR	14-SEP-2000	2000US-0233401
PR	14-SEP-2000	2000US-0233054
PR	14-SEP-2000	2000US-0233063
PR	14-SEP-2000	2000US-0233065
PR	14-SEP-2000	2000US-0233066
PR	21-SEP-2000	2000US-0234274
PR	21-SEP-2000	2000US-0234275
PR	25-SEP-2000	2000US-0234397
PR	25-SEP-2000	2000US-0234398
PR	25-SEP-2000	2000US-0234399
PR	26-SEP-2000	2000US-0235484
PR	27-SEP-2000	2000US-0235484
PR	27-SEP-2000	2000US-0235834
PR	27-SEP-2000	2000US-0235836
PR	29-SEP-2000	2000US-0236327
PR	29-SEP-2000	2000US-0236367
PR	29-SEP-2000	2000US-0236368
PR	29-SEP-2000	2000US-0236369
PR	29-SEP-2000	2000US-0236370
PR	29-SEP-2000	2000US-0236371
PR	29-SEP-2000	2000US-0236372
PR	02-OCT-2000	2000US-0237033
PR	02-OCT-2000	2000US-0237038
PR	02-OCT-2000	2000US-0237039
PR	02-OCT-2000	2000US-0237040
PR	02-OCT-2000	2000US-0237041
PR	02-OCT-2000	2000US-0239335
PR	13-OCT-2000	2000US-0239337
PR	13-OCT-2000	2000US-0239360
PR	20-OCT-2000	2000US-0240960
PR	20-OCT-2000	2000US-0241221
PR	20-OCT-2000	2000US-0241785
PR	20-OCT-2000	2000US-0241786
PR	20-OCT-2000	2000US-0241787
PR	20-OCT-2000	2000US-0241808
PR	20-OCT-2000	2000US-0241826
PR	20-OCT-2000	2000US-0244617
PR	01-NOV-2000	2000US-0244617
PR	08-NOV-2000	2000US-0244674
PR	08-NOV-2000	2000US-0244675
PR	08-NOV-2000	2000US-0244676
PR	08-NOV-2000	2000US-0244677
PR	08-NOV-2000	2000US-0244678
PR	08-NOV-2000	2000US-0244679
PR	08-NOV-2000	2000US-0245523
PR	08-NOV-2000	2000US-0245524
PR	08-NOV-2000	2000US-0245525
PR	08-NOV-2000	2000US-0245526
PR	08-NOV-2000	2000US-0245527
PR	08-NOV-2000	2000US-0245528
PR	08-NOV-2000	2000US-0245532
PR	08-NOV-2000	2000US-0246609
PR	08-NOV-2000	2000US-0246610
PR	08-NOV-2000	2000US-0246611
PR	08-NOV-2000	2000US-0246613
PR	17-NOV-2000	2000US-0249207
PR	17-NOV-2000	2000US-0249208
PR	17-NOV-2000	2000US-0249209
PR	17-NOV-2000	2000US-0249210
PR	17-NOV-2000	2000US-0249211
PR	17-NOV-2000	2000US-0249212
PR	17-NOV-2000	2000US-0249216
PR	17-NOV-2000	2000US-0249217
PR	17-NOV-2000	2000US-0249218
PR	17-NOV-2000	2000US-0249244
PR	17-NOV-2000	2000US-0249245
PR	17-NOV-2000	2000US-0249264
PR	17-NOV-2000	2000US-0249265
PR	17-NOV-2000	2000US-0249297
PR	17-NOV-2000	2000US-0249299
PR	17-NOV-2000	2000US-0249300
PR	17-NOV-2000	2000US-0249309

PR	01-DEC-2000;	2000US-0250160.
PR	01-DEC-2000;	2000US-0250391.
PR	05-DEC-2000;	2000US-0251030.
PR	05-DEC-2000;	2000US-0251968.
PR	05-DEC-2000;	2000US-0256719.
PR	06-DEC-2000;	2000US-0251479.
PR	08-DEC-2000;	2000US-0251856.
PR	08-DEC-2000;	2000US-0251866.
PR	08-DEC-2000;	2000US-0251869.
PR	08-DEC-2000;	2000US-0251988.
PR	08-DEC-2000;	2000US-0251990.
PR	11-DEC-2000;	2000US-0254097.
PR	05-JAN-2001;	2001US-0259678.
XX		
PA	(HUMA-)	HUMAN GENOME SCI INC.
XX		
PI	Rosen CA,	Barash SC, Ruben SM;
XX		
DR	WPI;	2001-488783/53.
XX		
DR	N-PSDB;	AAS26173.
XX		
PT	New nucleic acid molecules encoding	461 human secreted proteins for
PT	diagnosing, preventing, treating or	ameliorating medical conditions and
XX	used as food additives or preservatives	-
XX		
CS	Claim 11; SEQ ID No 1139;	980BP; English.

The invention relates to isolated nucleic acid molecules and their encoded secreted proteins. The nucleic acids and proteins are used to prevent, treat or ameliorate a medical condition in e.g. humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used in diagnosing a pathological condition or susceptibility to a pathological condition. Antibodies to the proteins can also be used in alleviating symptoms associated with the disorders and in diagnostic immunoassays e.g. radioimmunoassays or enzyme linked immunosorbent assays (ELISAs). Disorders which are diagnosed or treated include autoimmune diseases e.g. Rheumatoid arthritis, hyperproliferative disorders e.g. neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g. Alzheimer's disease, infections caused by bacteria, viruses and fungi and ocular disorders e.g. corneal infection, and many other disorders listed in the specification. The polypeptides can also be used to aid wound healing and epithelial cell proliferation, to prevent skin aging due to sunburn, to maintain organs before transplantation, for supporting cell culture of primary tissues, to regenerate tissues and in chemotaxis. The polypeptides can also be used as a food additive or preservative to increase or decrease storage capabilities, fat content, lipid, protein, carbohydrate, vitamins, minerals, cofactors and other nutritional components. The present sequence represents a novel secreted protein of the invention.

Query Match	3.7%;	Score 285;	DB 22;	Length 591;
Best Local Similarity	22.0%;	Pred. No. 5.1e-10;		
Matches 130;	Conservative 94;	Mismatches 250;	Indels 118;	Gaps 22;

```

OY 17 RRRKCEPQOLGSPKRNKRNRREONKRIEELAEIIFNFNDINENKPRKALIKET 76
Db 53 tRkrdsdsdpsgeahsqtekrtdkmmnlleelssami---pqcpmarkldkllvltma 108
OY 77 VKQIQIKOEKOEFAAANIDEVQKSVSSSTGQGVIRKDALGPMMLALDGFVFNLE-GN 135
Db 109 vqlhslstkgltstsyg-----snypsfldqnelrhllktaegflvsgcergk 158
OY 136 VVEFENATQVIRYNOEELMKNKSVYSILHVGHTPEVKNLPLKSLYNG-----GSMGEP 190
Db 159 lltvsksvskllyngqastlqgsldldflhpkrvak-vkeqslsfdisprekldaktqg 217
OY 191 PRRNSH-----TFNCRMLVYKPLPDSEERG-----HONOEAHOKRETMQC--FAVS 233
Db 218 vshnlhagrtlvysgsrrtsfcrllskckisvkeegcjlpsnkkkkehkytlthcgylls 277
OY 234 QPKSIKERGEDLQS-----CLICVARR---VPMKERPLPSPSESTFTTRDOLGKIT 281

```



[illegible][illegible]

```
Oy      815 NSQLPQLEPPDRPCAPASVOKALINDIMLOMTLHENSVPYVCAQKATLRALSOSTFENNR 874
Db      721 ----eqafqldtsygddppptssshlmwkrmkslmgylcptlmp----- 757

Oy      875 PQLGLRLLPNQNLPLDRILOSPTAGAPPPIPRNSPSYVIPPOCMNMGOGMIGNOG---- 930
Db      758 -----dkttisammapdefqkasmrgly--qpriihp-----pqbpsttsgenaktgfppq 807

Oy      931 -----ILGNSSTGMIGNSASRPRTMESEWAPQSSAVRYTCATTISANRRPYQG 978
Db      808 cyaaagfdqdygppgaqkvsgvasrlllpfsepyllpeltyrdc-----evnvpypg 857
```

Search completed: September 7, 2002, 10:38:07  
Job time: 319 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 7, 2002, 10:35:38 ; Search time 31.5 Seconds

(without alignments)  
4465.866 Million cell updates/sec

Title: US-09-842-256-2

Perfect score: 7631

Sequence: 1 MSGMGENTSDPSRAETRRKRRK.....NOLPGMDMIKQEGDTTRKYC 1464

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Listing first 45 summaries

1: pir1.\*

2: pir2.\*

3: pir3.\*

4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7216	94.6	1462	2 T42639	glucocorticoid rec
2	6659.5	87.3	1463	2 T30193	nuclear receptor c
3	2729	35.8	1424	2 T03851	thyroid hormone re
4	1033.5	13.5	1061	2 A57620	steroid receptor c
5	340	4.5	2232	2 T34434	hypothetical prote
6	319	4.2	1505	2 JC4851	hypoxia-inducible
7	313.5	4.1	4957	2 T03455	ALR protein - huma
8	313.5	4.1	5262	2 T03454	ALR protein - huma
9	309	4.0	2715	2 T13049	eyelid - fruit fly
10	297.5	3.9	1088	2 H96747	unknown protein T1
11	292	3.8	2649	2 T51023	hypothetical prote
12	287	3.8	1234	2 T30160	hypothetical prote
13	282	3.7	3498	2 T22330	hypothetical prote
14	280	3.7	925	2 T19361	hypothetical prote
15	277	3.6	2414	2 A54277	transcription adap
16	275	3.6	2722	2 T20532	hypothetical prote
17	273.5	3.6	1366	2 B86292	F7H2.12 protein -
18	272	3.6	626	2 JE0270	Antt-like PAS prot
19	270.5	3.5	2738	2 E88320	protein F07A11.6 l
20	270	3.5	667	2 JC7771	hypoxia inducible
21	270	3.5	1589	2 T13606	hypothetical prote
22	269.5	3.5	826	2 T13972	hypoxia-inducible
23	269	3.5	626	2 JC5405	she protein - musc
24	269	3.5	1212	2 T13804	hypothetical prote
25	269	3.5	1952	2 T48814	hypothetical prote
26	268.5	3.5	1058	2 T30556	aryl hydrocarbon r
27	268	3.5	1142	2 T00022	B120 protein - hum
28	266.5	3.5	2090	2 T52058	probable transform
29	264.5	3.5	1307	2 T25563	hypothetical prote

30	263.5	3.5	1059	2 T30557	aryl hydrocarbon r
31	261.5	3.4	2282	2 T42717	DNA-binding protei
32	258.5	3.4	3507	2 T34513	hypothetical prote
33	258	3.4	2215	2 T16871	hypothetical prote
34	257.5	3.4	1655	2 T13998	hypothetical prote
35	255.5	3.3	813	2 JC5809	gene mastermind pr
36	254.5	3.3	2606	2 T24157	hypoxia-inducible
37	253.5	3.3	2584	2 T24158	hypothetical prote
38	252	3.3	1819	2 T32008	hypothetical prote
39	252	3.3	2441	2 S39161	hypothetical prote
40	252	3.3	3190	2 T13828	CREB-binding prote
41	251	3.3	1794	2 T38459	hypothetical diver
42	250	3.3	1023	2 T13068	CLOCK protein - fir
43	250	3.3	1596	2 A33106	neurogenic locus m
44	248.5	3.3	3968	2 A44265	trithorax homolog
45	248	3.2	2175	1 S03170	homeotic protein c

#### ALIGNMENTS

RESULT	1
T42639	
glucocorticoid receptor interacting protein GRIP1 - mouse	
C:Species: Mus musculus (house mouse)	
C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 11-Jan-2000	
C:Accession: T42639	
R:Hong, H.; Kohli, K.; Garabedian, M.J.; Stallcup, M.R.	
Mol. Cell. Biol. 17, 2735-2744, 1997	
A>Title: GRIP1, a transcriptional coactivator for the AF-2 transactivation domain of	
A:Reference number: Z22229; MUID:97265407	
A:Accession: T42639	
A>Status: preliminary; translated from GB/EMBL/DBJ	
A:Molecule type: mRNA	
A:Residues: 1-1462 <HON>	
A:Cross-references: EMBL.U39060; NID:G1853979; PID:G1853980; PIDN:AAC53151.1	
C:Genetics:	
A:Gene: GRIP1	
C:Function:	
A:Description: transcriptional co-activator for steroid receptors and nuclear recepto	
A>Note: AF-2 requires GRIP1 as a coactivator, but AF-1 does not	

Query Match	94.6%	Score 7216;	DB 2;	Length 1462;
Best Local Similarity	94.1%	Pred. No. 0;		
Matches 1377;	Conservative 43;	Mismatches 42;	Indels 2;	Gaps 1;
QY	1	MSGMGENTSDPSRAETRRKKECPDOLGSPKRNTEKRNREQENKYTEIETAEIIFANFNDI	60	
DB	1	MSGMGENTSDPSRAETRRKKECPDOLGSPKRNTEKRNREQENKYTEIETAEIIFANFNDI	60	
QY	61	DNFNKPKDKCALIKETVKQIQIRKQEKAAANIDEVOKSDVSSFGGCVIDKDALGPMML	120	
DB	61	DNFNKPKDKCALIKETVKQIQIRKQEKAAANIDEVOKSDVSSFGGCVIDKDALGPMML	120	
QY	121	EALDGFPPVNLKGNVWVPESENVTOYLRNOBELMKNKSVSYLTHVGHDTPEFKNLLPESI	180	
DB	121	EALDGFPPVNLKGNVWVPESENVTOYLRNOBELMKNKSVSYLTHVGHDTPEFKNLLPESI	180	
QY	181	VNGSGMSGEPFRNRSHTEFCRMLVPLPDSEEGHDNOEAHQKYETMOCFAVSOQPSIKE	240	
DB	181	VNGSGMSGEPFRNRSHTEFCRMLVPLPDSEEGHDNOEAHQKYETMOCFAVSOQPSIKE	240	
QY	241	EGEDLOSLICVARRVPKKEKREPVLPSSSEFTTRDLOGKITSLDSTFMRAMKPEMEDLV	300	
DB	241	EGEDLOSLICVARRVPKKEKREPVLPSSSEFTTRDLOGKITSLDSTFMRAMKPEMEDLV	300	
QY	301	RRCTOKPFAOHGSESVSAKKRHNEVLRQAFSQTYPFSSDGLVAQAOKRSKILRSOT	360	
DB	301	RRCTOKPFAOHGSESVSAKKRHNEVLRQAFSQTYPFSSDGLVAQAOKRSKILRSOT	360	
QY	361	TNEPOLVLSLHMLHREQWCVANPDLTGOTMGKPLNPITSSNSPAHQALCSGPNPGDMTIS	420	
DB	361	TNEPOLVLSLHMLHREQWCVANPDLTGOTMGKPLNPITSSNSPAHQALCSGPNPGDMTIS	420	

Db 361 TNEPQVYISLHMLHREQNVCMNPDLTGQAMGRPLNPISSSSPAHQALCSGNPQDMTLG 420  
 QY 421 SNINFPINGKEQMGMPGRFGSGGMNHYSGMOATTPQGSNNVALKKNNSPQSSPGMNP 480  
 Db 421 SNINFPINGKEQMGMPGRFGSGGMNHYSGMOATTPQGSNNVALKKNNSPQSSPGMNP 480  
 QY 481 QPMSLSPRRHMSGVAGSPRIIPSPQSPAGSLHSPVGVCSSTGNSHSTYNSSLNALQAL 540  
 Db 481 QASSVLSPPRRMSPGVAGSPRIIPSPQSPAGSLHSPVGVCSSTGNSHSTYNSSLNALQAL 540  
 QY 541 SEGHGVSLSGLSLASPDLMKGNLQNSPVNMNPPPLSKMGLSDSKDCCFLYGPSPGCTGOA 600  
 Db 541 SEGHGVSLSGLSLASPDLMKGNLQNSPVNMNPPPLSKMGLSDSKDCCFLYGPSPGCTGOA 600  
 QY 601 ESSCHPEQKEETNDPNLPVAVSERADGSRSLHDSKQOTKLLQLLTTKSDQMEPSPLASS 660  
 Db 601 EASCHPEQKPNDSMPQAASGDRAGHSLHDSKQOTKLLQLLTTKSDQMEPSPLASS 660  
 QY 661 LSTNDKSTGSLSGSGSTHGTSLAKEKHKILHRLLDSSSPVDLAKLTAEATGKDLQESS 720  
 Db 661 LSTNDKSTGSLSGSGSTHGTSLAKEKHKILHRLLDSSSPVDLAKLTAEATGKDLQESS 720  
 QY 721 STAPGSEVTIKQEPVSKKKNALLRYLLDKDPTKDIGLPETPKLERLDSKTDPAANTK 780  
 Db 721 STAPGSEVTIKQEPVSKKKNALLRYLLDKDPTKDIGLPETPKLERLDSKTDPAANTK 780  
 QY 781 LIAMKTEKEEMSEEPDQPGSELNLEILLDLQNSQLPOLFPDTPRGAPABSVKQAI 840  
 Db 781 LIAMKTEKEEMSEEPDQPGSELNLEILLDLQNSQLPOLFPDTPRGAPABSVKQAI 840  
 QY 841 NDLMQTLAENSPTVPAQKLTALRISQSTFNPNRPGQLGRLPNQNLPLDITTLQSPGAG 900  
 Db 841 NDLMQTLADSSVPVPAQAALMSQSTFNPNRPGQLGRLPNQNLPLDITTLQSPGAG 900  
 QY 901 PEPPIRNSSTSYVLPQPGMNGNOGMIGNQNLGSSSTGMIGNSARPTMPSGEMAPQSSA 960  
 Db 901 PEPPIRNSSTSYVLPQPGMNGNOGMIGNQNLGSSSTGMIGNSARPTMPSGEMAPQSSA 960  
 QY 961 VRVTCATTSAMNRPVQGMIRNPAASIPMRPSSQPGQROTLOQVYNNIGPSELEMMNG 1020  
 Db 961 VRVTCATTSAMNRPVQGMIRNPAASIPMRPSSQPGQROTLOQVYNNIGPSELEMMNG 1020  
 QY 1021 PQVSQQAAPRNOTAPWPESTLPTDQASFASQONQPPSSSDDLCPHRAESPSDECAL 1080  
 Db 1021 PQVSQQAAPRNOTAPWPESTLPTDQASFASQONQPPSSSDDLCPHRAESPSDECAL 1080  
 QY 1081 DOYLALRNFDGLLEIDRALGIPELVSOQAVPPEQFSQSDSNMLBOKAPVFPQOYASQ 1140  
 Db 1081 DOYLALRNFDGLLEIDRALGIPELVSOQAVPPEQFSQSDSNMLBOKAPVFPQOYASQ 1140  
 QY 1141 AQMAQGSYSFPMQDPNFTHTMGQRPSTATLRMQPRPGLRPTGLVQNPQRLQLOHRLQAO 1200  
 Db 1141 AQMAQGSYSFPMQDPNFTHTMGQRPSTATLRMQPRPGLRPTGLVQNPQRLQLOHRLQAO 1200  
 QY 1201 QNTQPLMNQSNVSNVLTLRGVPYTPQAPINAOMLAORQELLNQHRLQOMHQOQOYQ 1260  
 Db 1201 QNTQPLMNQSNVSNVLTLRGVPYTPQAPINAOMLAORQELLNQHRLQOMHQOQOYQ 1260  
 QY 1261 RTLLMNRQGLNMTPSMVAAPSPATMSNPRIPOANAQOFPFPYNYGISQOPDGFGTGAT 1320  
 Db 1261 RTLLMNRQGLNMTPSMVAAPSPATMSNPRIPOANAQOFPFPYNYGISQOPDGFGTGAT 1320  
 QY 1321 POSPLMSPRAHQPSPMAQSOQANPAYQAPSDINGMAGNMGNMSMSQSPPHFGQAN 1380  
 Db 1321 POSPLMSPRAHQPSPMAQSOQANPAYQAPSDINGMAGNMGNMSMSQSPPHFGQAN 1380  
 QY 1381 TSMTYSNNMNTINVSMAINTGSMSSNMQNTGOISMTSVTSVTSGLSKMGPQVNDPALRG 1440  
 Db 1381 TSMTYSNNMNTINVSMAINTGSMSSNMQNTGOISMTSVTSVTSGLSKMGPQVNDPALRG 1440  
 QY 1441 NLFPNQLPGMDMTKQSGDPTTKYC 1464  
 Db 1441 NLFPNQLPGMDMTKQSGDPTTKYC 1464  
 QY 1439 NLFPNQLPGMDMTKQSGDPTTKYC 1462  
 Db 1439 NLFPNQLPGMDMTKQSGDPTTKYC 1462

RESULT 2  
 T30193  
 C:Species: Mus musculus (house mouse)  
 C:Date: 22-Oct-1999 #sequence\_rev1510 22-Oct-1999 #text\_change 22-Oct-1999  
 R:Torchiya, J.; Rose, D.W.; Inostroza, J.; Kamel, Y.; Westlin, S.; Glass, C.K.; Rosenfe  
 Nature 387, 677-684, 1997  
 A:Title: The transcriptional co-activator p/CIP binds CBP and mediates nuclear-recept  
 A:Reference number: Z20768; MUID:97336097  
 A:Accession: T30193  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-1463 <TOR>  
 A:Cross-references: EMBL:AF000582; NID:92213816; PTD:92213817; PIDN:AA61575.1

Query Match 87.3%; Score 6659.5; DB 2; Length 1463;  
 Best Local Similarity 87.6%; Pred. No. 0;  
 Matches 1298; Conservative 60; Mismatches 87; Indels 37; Gaps 10;

QY 1 MSGMGENTSPPSAEFRKREKCPDQGLSPKRNTERKRNQENKYTEELAFANFNDI 60  
 Db 1 MSGMGENTSPPSAEFRKREKCPDQGLSPKRNTERKRNQENKYTEELAFANFNDI 60  
 QY 61 DNEFKRPDKCAILKETVQJROIKEOEKAAANIDEVQKSDVSTGGVTDKALGPMML 120  
 Db 61 DNEFKRPDKCAILKETVQJROIKEOEKAAANIDEVQKSDVSTGGVTDKALGPMML 120  
 QY 121 EALDGEFFVNLGNNVVESENVTQYLRYNOBELMKNKSVYSLHVGDTHEVKNLLPKST 180  
 Db 121 EALDGEFFVNLGNNVVESENVTQYLRYNOBELMKNKSVYSLHVGDTHEVKNLLPKST 180  
 QY 181 VNGGSMGGEPRRNSHTFNCRMLYKPLPDBEECHDQNEHQKYTEYQCAVAPKSIKE 240  
 Db 181 VNGGSMGGEPRRNSHTFNCRMLYKPLPDBEECHDQNEHQKYTEYQCAVAPKSIKE 240  
 QY 241 EGEDLQSLCLVARRVPMKER-----PVLPSSESFTTRDQGLKITSIDPTSM 288  
 Db 241 EGEDLQSLCLVARRVPMKER-----PVLPSSESFTTRDQGLKITSIDPTSM 288  
 QY 289 RAAM-KPWEDLVRCIQKFNQHEGESYAKRHNNHEVLROGLAFSQIYRFLSDGTIV 347  
 Db 289 RAAM-KPWEDLVRCIQKFNQHEGESYAKRHNNHEVLROGLAFSQIYRFLSDGTIV 347  
 QY 348 AAQTKSLIRSQTTNEPQVLYISLHMLHREQNVCMNPDLTGQAMGRPLNPISSSPAHQA 407  
 Db 348 AAQTKSLIRSQTTNEPQVLYISLHMLHREQNVCMNPDLTGQAMGRPLNPISSSPAHQA 407  
 QY 408 LCGSNPQODMTLSSNINFPINGKEQMGMPGRFGSGGMNHYSGMOATTPQGSNNVALK 467  
 Db 408 LCGSNPQODMTLSSNINFPINGKEQMGMPGRFGSGGMNHYSGMOATTPQGSNNVALK 467  
 QY 468 NSPQSSPGMNPQASVYSLPQRMSPGVAGSPRIIPSPQSPAGSLHSPVGVCSSTGNSH 527  
 Db 468 NSPQSSPGMNPQASVYSLPQRMSPGVAGSPRIIPSPQSPAGSLHSPVGVCSSTGNSH 527  
 QY 528 STYNSSLNALQALSEGHGVSLSGLSLASPDLMKGNLQNSPVNMNPPPLSKMGLSDSK 587  
 Db 528 STYNSSLNALQALSEGHGVSLSGLSLASPDLMKGNLQNSPVNMNPPPLSKMGLSDSK 587  
 QY 588 LYGEPSECTGQAESSCHPEQKETNDPNLPVAVSERADGSRSLHDSKQOTKLLQLLTT 647  
 Db 588 LYGEPSECTGQAESSCHPEQKETNDPNLPVAVSERADGSRSLHDSKQOTKLLQLLTT 647  
 QY 648 KSDQMEPSPLASSLSDTNKSTGSLPGSGSTHGTSLAKEKHKILHRLLDSSSPVDLAKLT 707  
 Db 648 KSDQMEPSPLASSLSDTNKSTGSLPGSGSTHGTSLAKEKHKILHRLLDSSSPVDLAKLT 707  
 QY 708 AEATGKDLQESSSTAPGSEVTIKQEPVSKKKNALLRYLLDKDPTKDIGLPETPKLE 767  
 Db 708 AEATGKDLQESSSTAPGSEVTIKQEPVSKKKNALLRYLLDKDPTKDIGLPETPKLE 767

Db 708 AEATGKELSOESSSTAPEGSEVTVKOEAPSPKKENALLRLLDKDITKIGLPEITPKLE 767  
 Oy 768 RLDSKTDPASTMTKIAMTKEKEMSEFEPDQPGSELNDLEETLDDQNSQLPOLPEPDRP 827  
 Db 768 RLDSKTDPASTMTKIAMTKEEVSFEESDQPGSELNDLEETLDDQNSQLPOLPEPDRP 827  
 Oy 828 GAPAGSYVDKQAIINDLMQITLTAENSPFVPGAKTALRISQSTFNNPRGQGLRLLPNQNL 887  
 Db 828 GAPAGSYVDKQAIINDLMQITLTAENSPFVPGAKTALRISQSTFNNPRGQGLRLLPNQNL 887  
 Oy 888 PLDITLLOSPGTAGPFPPLRNSPSPVIRPQPMGNOGMITGNOGLGNSSTGMIGNSASRP 947  
 Db 888 PLDITLLOSPGTAGPFPPLRNSPSPVIRPQPMGNOGMITGNOGLGNSSTGMIGNSASRP 947  
 Oy 948 TMRPGEMAPSSAVRVTCAT--TSAMNRPVGGMR--NPASISMRPSSQPGQRTL 1002  
 Db 948 TMRPGEMAPSSAVRVTCAT--TSAMNRPVGGMR--NPASISMRPSSQPGQRTL 1002  
 Oy 1003 OSQVMNIGPSELNMGSPQYSQOQAPNPQAPMPESILPIDQASFAQNRQPGSSPD 1062  
 Db 1004 OSQVMNIGPSELNMGSPQYSQOQAPNPQAPMPESILPIDQASFAQNRQPGSSPD 1063  
 Oy 1063 LLCFHPAAESPDSGALIDQLYLALRNFDELIDRALGIDELVSQSAVDPEQSSDS 1122  
 Db 1064 LLCFHPAAESPDSGALIDQLYLALRNFDELIDRALGIDELVSQSAVDPEQSSDS 1123  
 Oy 1123 NIMLEOKAPVPPOOYASQOQAGSYSPMDPNFTMGORSPYATLRMOPRGLRPGLY 1182  
 Db 1124 NIMLEOKAPVPPOOYASQOQAGSYSPMDPNFTMGORSPYATLRMOPRGLRPGLY 1183  
 Oy 1183 QNQPNOQLRLQHRQLAQOQNRPLMNOISVSNVNLTLRPGVPTQAPINAOMLAQOREI 1242  
 Db 1184 QNQPNOQLRLQHRQLAQOQNRPLMNOISVSNVNLTLRPGVPTQAPINAOMLAQOREI 1243  
 Oy 1243 LNOHLRQRHMQOQOQOQRTLMAGCGIANTPSSVAVSGAMATSNRPITPANAQOFPFP 1302  
 Db 1244 LNOHLRQRHMQOQOQOQRTLMAGCGIANTPSSVAVSGAMATSNRPITPANAQOFPFP 1301  
 Oy 1303 PNYSIGSOQDPGFTGATTPQSPFLMSPRRAHTQSPMGOQSOANPAYOAFSDINGNAQMG 1362  
 Db 1302 PNYSIGSOQDPGFTGATTPQSPFLMSPRRAHTQSPMGOQSOANPAYOAFSDINGNAQMG 1361  
 Oy 1363 GNSMFQSOQSPHFGQOANTSMYSNNMNIIVSMATNTGSMSSNMOTGOISMTSVTS 1422  
 Db 1362 GNSMFQSOQSPHFGQOANTSMYSNNMNIIVSMATNTGSMSSNMOTGOISMTSVTS 1421  
 Oy 1423 GLSSMGPEOVNDPALRGCNLFPPNOLPGMDMTKQSGDTTRKTC 1464  
 Db 1422 GLSSMGPEOVNDPALRGCNLFPPNOLPGMDMTKQSGDTTRKTC 1463  
 RESULT 3  
 T03851  
 thyroid hormone receptor activator molecule - human  
 C:Species: Homo sapiens (man)  
 C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 01-Dec-2000  
 C:Accession: T03851; T03749; T03443  
 R:Tekeshita, A.; Cardona, G.R.; Koibuchi, N.; Suen, C.S.; Chin, W.W.  
 J. Biol. Chem. 272, 27629-27634, 1997  
 A:Title: TRAM-1, a novel 160-kDa thyroid hormone receptor activator molecule, exhibits d  
 A:Reference number: 215120; M01D:98010595  
 A:Accession: T03851  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-1424 <TRAM>  
 A:Cross-references: EMBL:AF016031; NID:g2584879; PIDN:AA051849.1; PID:g2584880  
 R:Anzick, S.L.; Kononen, J.; Walker, R.L.; Azorsa, D.O.; Tanner, M.M.; Guan, X.Y.; Sauter  
 Science 277, 965-968, 1997  
 A:Title: AIB1, a steroid receptor coactivator amplified in breast and ovarian Cancer.  
 A:Reference number: 215053; M01D:97400625  
 A:Accession: T03749  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA

A:Residues: 1-1213,1218-1424 <ANZ>  
 A:Cross-references: EMBL:AF012108; NID:g2331249; PIDN:AA051677.1; PID:g2331250  
 R:Li, H.; Gomes, P.J.; Chen, J.D.  
 Proc. Natl. Acad. Sci. U.S.A. 94, 8479-8484, 1997  
 A:Title: RAC3, a steroid/nuclear receptor-associated coactivator that is related to S  
 A:Reference number: 214950; M01D:97385128  
 A:Accession: T03443  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-130,'EA',133-1213,1218-1273,1277-1424 <LIH>  
 A:Cross-references: EMBL:AF010227; NID:g2318005; PIDN:AA051663.1; PID:g2318006  
 A:Experimental source: cell line HeLa  
 C:Genetics:  
 A:Gene: TRAM-1; AIB1; RAC3  
 A:Map position: 20  
 Query Match 35.8%; Score 2729; DB 2; Length 1424;  
 Best Local Similarity 42.5%; Pred. No. 1.4e-127;  
 Matches 644; Conservative 243; Mismatches 451; Indels 176; Gaps 44;  
 Oy 1 MSGEENTSDPSRAETRRKRCQPOLGSPKRTNTERKRNDEKRYIEELAFRANFNDI 60  
 Db 1 MSGLENTL-DPLASDSRRKRLPCDTPGGLTCSEKRRREDESKYIEELAFRANFNDI 59  
 Oy 61 DNFNPKPKCALKEVYQAIQRIQKEKAAANIDEVOKSPVSSGTGGVTDKDLGPMML 120  
 Db 60 DNFNPKPKCALKEVYQAIQRIQKEQK-TISNDDDVQKAVSVSTGGVTDKDLGPMML 118  
 Oy 121 EALDGFPEFVNLGCVNVFSENVATQYLRVNOELMKNVSILVHGDTPEFVKMLPKSI 180  
 Db 119 QALDGFLEFVNRGDNIVFSENVATQYLYQKQEDLVNVSIVILHEBRKDFLKN-LPKST 177  
 Oy 181 VNGSMGSEPPRRKSHFTNCMLYKPLPDESEBDHNOEAHQKRYEMOCRAVSPKSIKE 240  
 Db 178 VNGSVWINEVTRKSHFTNCMLKTPDILIEDINASEPMORRYETMOCFALSOPRAME 237  
 Oy 241 EGEDLOSCLICVARRVPRKERVLPSSSEFTTRDLOGKITSLTPRMAKRGWEDLY 300  
 Db 238 EGEDLOSCLICVARRVPRKERVLPSSSEFTTRDLOGKITSLTPRMAKRGWEDLY 297  
 Oy 301 RRCIOKFAHQEGESVSYAKRHHEVLRQGLAFQIYRFLSDGTLYAAQTKSLIRSQ 360  
 Db 298 RRCIOKFAHQEGESVSYAKRHHEVLRQGLAFQIYRFLSDGTLYAAQTKSLIRSQ 356  
 Oy 361 TNEPOLYISLHMLHREOVNVMNPDLTGQTMGKFLNPISSPAHQALCSGNPQDMTSL 420  
 Db 357 TNEPOLYISLHMLHREOVNVMNPDLTGQTMGKFLNPISSPAHQALCSGNPQDMTSL 412  
 Oy 421 SNINPPIGPKQEGMGPMGRGSGGMNHS---GMQA-TTPQGSNYALKNNSPQSSPG 476  
 Db 413 SNINPPIGPKQEGMGPMGRGSGGMNHS---GMQA-TTPQGSNYALKNNSPQSSPG 472  
 Oy 477 MNPQOPTSMSPRRHMSPGVAGSPRIPPOSPAGSLHSPGVCSSTGNSHSTYNSSLNA 536  
 Db 473 MNPQOPTSMSPRRHMSPGVAGSPRIPPOSPAGSLHSPGVCSSTGNSHSTYNSSLNA 525  
 Oy 537 LQALSEGHSVLSGLASPLDKMGNLONSPVNNMPPPLSKMSLSDSCDEFLYE--PSE 594  
 Db 526 LQALSEGHSVLSGLASPLDKMGNLONSPVNNMPPPLSKMSLSDSCDEFLYE--PSE 581  
 Oy 595 GTTQOASSCHPGQKFNNDNPLPVAVSSERADQSLRHDSKQGTKLQLLTTSDD--- 650  
 Db 582 GTTQOASSCHPGQKFNNDNPLPVAVSSERADQSLRHDSKQGTKLQLLTTSDD--- 634  
 Oy 651 -QMPSPPLASSL-----SDTNKDSGLSLGSGSTHSTSLKEKHLRLRLQDSSS 699  
 Db 635 -QMPSPPLASSL-----SDTNKDSGLSLGSGSTHSTSLKEKHLRLRLQDSSS 694  
 Oy 700 PVDIAKTLAETGKDLQSOESSSTAPEGSEVTVKOEAPSPKKENALLRLLDKDITKIG 758  
 Db 695 PVDIAKTLAETGKDLQSOESSSTAPEGSEVTVKOEAPSPKKENALLRLLDKDITKIG 751

**RESULT**

stereoid receptor coactivator 1 - human  
A57620

C.Species: Homo sapiens (man)  
C.Date: 08-Feb-1996 #sequence\_revision 08-Feb-1996 #text\_change 01-Dec-2000  
C.Accession: A57620; PC4362  
R.Onate, S.A.; Tsai, S.Y.; Tsai, M.J.; O'Malley, B.W.  
Science 270, 1354-1357, 1995  
A.Title: Sequence and characterization of a coactivator for the steroid hormone receptor  
A.Reference number: A57620; MUID:96089563  
A.Accession: A57620

A.Status: preliminary; not compared with conceptual translation  
A.Molecule type: mRNA  
A.Residues: 1-1061 <N>A>  
A.Cross-references: GB:U04936; NID:g1117914; PIND:AAC50305.1; PID:g1117915  
R.Hayashi, Y.; Ohnori, S.; Ito, T.; Seo, H.  
Biochem. Biophys. Res. Commun. 236, 83-87, 1997  
A>Title: A splicing variant of steroid receptor coactivator-1 (SRC-1E): The major isoform  
A.Reference number: PC4362; MUID:97366601  
A.Accession: PC4362

Detailed description of the sequence alignment results, showing coordinates and sequence details for various entries.

A:Molecule type: DNA  
A:Residues: 990-1061 <HAY>  
C:Comment: This protein interacts with nuclear receptor only when they are bound to the DNA  
C:Keywords: steroid hormone receptor

Query Match 13.5%; Score 1033.5; DB 2; Length 1061;  
Best Local Similarity 28.8%; Pred. No. 1,36-43;  
Matches 348; Conservative 175; Mismatches 335; Indels 349; Gaps 52;

QY 395 LMPISNSPAHQAALCSG--PGQDMLSSNIFPLNGKEQ-----MGMPGRF 441  
DB 10 YNP--SISPAHVARASTLPSPSNMNVSTRIN-----RQSSDLHSSHSNNSNSQSPF 61  
QY 442 GSGGGMNHYSGM-----OATPQGSNVAALKMSPSQSSPGMNPQGPMTSLPRHMGPGV 496  
DB 62 GCSPPSQIYANVALKKGAASS-QSSKPSLNLNPNMEGTGSLAQ---FMSPRQVYISGL 117  
QY 497 AGSPRIIPSPQFSP-AGSLHSPVGCSS--TGNSHSYTNSSLNALQALSEGHVSLGSSLA 553  
DB 118 ATRPRHNPNSFPPISTLSPSGMTSSACNNNNRSTSNIPYTSLOGMEGPNNSGFSAS 177  
QY 554 SPDLKMGNLQNSPVMNPNRPLSKMGSLSKSCFGLYGERPSEGTQQAASSCHPQOKT- 612  
DB 178 SPVLKQSSQNSPSRLNLP-AKAESKDKKELASTLANMT-----QSNSSSDGCPALDSG 231  
QY 613 ---NPNLIPPAVSSERADQOSRLHDSKGOTKLLQTLTKSD-QMESPGLASSLD----- 663  
DB 232 LLAHNDRL-----SGDGSKY--SQTSKIKYQLLTTTAQQLRHADIDISCKDVLST 281  
QY 664 --TNKDSGSLGSGSYTGTSLSKEKHKILHRLLDSSPVDLAKTLAATGKDLSSQSSS 721  
DB 282 GTSNAGSANSSSGSGSPSSHSLTAHKKILHRLDQ--GSPSDITTLSTVSPDKKD--SASTV 339  
QY 722 TAPG---SEVTLKQPVAPPKKE---NALRYLRLDKD---TKDILRETLRETLRD 770  
DB 340 SVTGVQGNSSSKLELDASKKESKDHQRLRLDKDEKDLRSTPNLSLDVKKYKVEKE 399  
QY 771 SKTDPAASNTKLIAMTKEKEEMSEFEPDQPSGSELDLMEILDLQNL-SQLPOLFPDTRPGA 829  
DB 400 -QMDPCNTNPTPMTKATPEEIKLEAQSFADLDQFDQLPTLEKRAQLPGICETDRMD- 457  
QY 830 PAGSYDKQAIIIDMLQILAENSPTVPVGAQKTALRIISSTFNPNRPGQLGRLLPNQNLPL 889  
DB 458 --GAATSVYTIKEI-----L 470  
QY 890 DITLQSPTGAGFPPIRNSSPYSYVIPQPGMGNQGMIGNQGLNMGSTGMIGNSASRPTM 949  
DB 471 PASLQSAT-----ARP-- 481  
QY 950 PSGEMAPQSSAVRYTCAATTSAMNRPVQGMATRNPAASIPMRPSSQPORQTLQSQVANI 1009  
DB 482 -----TSRLNR-----LP----- 489  
QY 1010 GPELEMMNGGPGYSSQQAAPRQGTAPWP-ESILPDIQASPAQONRQPGSSPDLLCEHP 1068  
DB 490 ---ELELELDNQFQPGP--GDQIPMTNNTYALNQ---SKSEDDQCSQLDELCPPT 541  
QY 1069 AAEPSDEGALLDQ--YIALRNFDELEIDRALGIPELVSSQAVD--PQGFSSQDSN- 1123  
DB 542 TVEGNEDEKALLEQVLSFGKDETELELDLDALEIDKTL-OGGLDLVLSERFFPQQAATP 600  
QY 1124 -IMLEOKAPVEFGQYASQAQMGQSYSPMODNFTMGORPSYATLRMOPR----- 1174  
DB 601 PLIMERRPNLVISOYSSPEPTAN-----LPSPFGQMVAKQPSLGTMPVQVTPPRKAFSPG 655  
QY 1175 -GLRPTGLVQN---QPNQLRLQLQRLQAQ-----QNRQPLNLQNSVSNVLTLRPG- 1223  
DB 656 MEMQPRQTLNRPAPRNQRLQDLQRLQGLQQLLHQNRQALINQPAATAPVGINRRSGMQ 715  
QY 1224 --VPIQAPIINAQMLAQREILLNQLRQRQMHQQAQVQQRILMMAGQ--GLAMPTSMYAP 1279  
DB 716 QQITTPRPPLNAQMLAQREILVSOQHRQRLQ---QQRALIMQAQSGFNGLNPS--- 767

OY	1280	SGAATMSNFRITQANAACQPPPPPPATGSGQODPGFTG----	ATTPOSSL-----	1325
Db	768	SGLEPVQGTNBRLLPQAGAPQGFPPPPNGTNPPTASPSQSLAANPEASLANRSMVSR	827	
OY	1326	-----MSPRMAHT-----	OSPM 1337	
Db	828	GMTGNIGGQGTCTININQMOQNVQYPGAGNVPQGEANFAPLSPESSNVPPIPPQSSL	887	
OY	1338	MOOSQANPAYOAPSDINGNAQCMGNSMFQ--OSPHEGOQANTSMYNNNNINVSMA	1395	
Db	888	LOOTPPASQSP-DKKAQOQGAIGNNVFQSAVONP---TPAPQGYI-NMSTIVSMA	942	
OY	1396	TNTGMSNMOMQGTISMTSVTSVTSGLSSPPOVNDPALRGENTLPNOLPGDMTKQ	1455	
Db	943	GGTNTVQNNPPMAAQOM---SLQAPGNTVCPEDQINDPALRTGLCYCNQSLSDLLKT	999	
OY	1456	EGDTTRK 1462		
Db	1000	EADGTQD 1006		

RESULT 5  
T34434  
hypothetical protein K06A9.1a - *Caenorhabditis elegans*  
C:Species: *Caenorhabditis elegans*  
C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 18-Feb-2000

A;Description: The sequence of C. elegans cosmid K06A9 submitted to the EMBL Data Library, December 1996

A; Reference number: Z21525

A:Accession: T34434  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-2232 <GEI>  
A:Cross-references: EMBL:08046, PIDN:AAC70890.1; GSPDB:GM00028; CESP:K06A9.1a  
A:Experimental source: strain Bristol N2; clone K06A9

A;Gene: CESP:K06A9.1a

A; Introns: 38/1; 75/3; 103/3; 132/2; 158/2; 222/1; 1088/1; 1367/1; 2039/1; 2049/1; 2075/1

Query Match	4.58;	Score 340;	DB 2;	Length 2232;
Best Local Similarity	20.88;	Pred. No. 1e-08;		
Matches 258;	Conservative 152;	Mismatches 489;	Indels 344;	Gaps 51;

QY 387 TGCTMGKPLNPISNSPAHQALCSNP-----GQDMFLSSNINFPINCPKQMGMPMG 439  
 |||: : ||| ||| : ||| | :||: : : :  
 Db 515 TGTSTV--TVVPGSSSTSPAPSS--SPNDSSTPASTGTITIGSSSIIVS-----TVSGS 564

QY      440    RCGSGGMNHHVSGMQATTPOGSNYALKMNS---PSQSSPGMPGQPTSMLSPRHRKSPGV    496  
         ||| : : : ||| : : : ||| | | | | |  
Db     565    TWSGTGTSTOSTLASSTATPSSSSTWPSSSSPQSSQAPAPNTGGTTP--SOTSOSP--    620

Df 621 --SPSMNDSSTPGSSOSTITPEGSTASSPFGSTGSEFVATEETTSOSTYPSGSSIGT- 677

Qy 497 AGSPRIIPSQFSACSLHPVCVCSTGNHSHS-YTNSSLNALQLAEGHGVSIGSSLASP 555

QY 556 DLKMGNTLONSPPVNNMPPLSKMGSLDSKDCEGLYGEPSEGTGGAESSCHPEGEKETNDP 615  
:  
db 678 -OSTNSBPSPSSLS-PTSGSMSTLS-----EPPSSST---OSS---GAOSITLTP 720

```
OY      616 NLPVAVSSERAD-----GQSR L HDSK G--QT K L I Q L L T T K S D O M P S P L A S S I D T P N K D S T   669
```

670 GSPDGGSTH--GTSLEKEHKHILHRLDSSSPVDLAKITAEATGKDLQESSSTAPGSE 727

0y 728 VTIKQEPVS-----PKKENALLRYLLDKDPTKDIGHPEITPKLERLDSKTDp 775

Db 828 VYVGSTEAETSGSSVASSSPAPSTQN-----NPSTSGSSMITQSPYP 872

```

      : : : | | | : |
873 SOSTPVESTTP-----SPG-SPGYL-----TSTSPSPQSTTGSTQGST 914
Db
```

DB	915	SPGISTSEEMTSQSGTQTPGSTGTVTQPTSVSDSTSSGTVTVGSGTEGSSSDIPSTQ	974
01	---	---	---
02	---	---	---
03	---	---	---
04	---	---	---
05	---	---	---
06	---	---	---
07	---	---	---
08	---	---	---
09	---	---	---
10	---	---	---
11	---	---	---
12	---	---	---
13	---	---	---
14	---	---	---
15	---	---	---
16	---	---	---
17	---	---	---
18	---	---	---
19	---	---	---
20	---	---	---
21	---	---	---
22	---	---	---
23	---	---	---
24	---	---	---
25	---	---	---
26	---	---	---
27	---	---	---
28	---	---	---
29	---	---	---
30	---	---	---
31	---	---	---
32	---	---	---
33	---	---	---
34	---	---	---
35	---	---	---
36	---	---	---
37	---	---	---
38	---	---	---
39	---	---	---
40	---	---	---
41	---	---	---
42	---	---	---
43	---	---	---
44	---	---	---
45	---	---	---
46	---	---	---
47	---	---	---
48	---	---	---
49	---	---	---
50	---	---	---
51	---	---	---
52	---	---	---
53	---	---	---
54	---	---	---
55	---	---	---
56	---	---	---
57	---	---	---
58	---	---	---
59	---	---	---
60	---	---	---
61	---	---	---
62	---	---	---
63	---	---	---
64	---	---	---
65	---	---	---
66	---	---	---
67	---	---	---
68	---	---	---
69	---	---	---
70	---	---	---
71	---	---	---
72	---	---	---
73	---	---	---
74	---	---	---
75	---	---	---
76	---	---	---
77	---	---	---
78	---	---	---
79	---	---	---
80	---	---	---
81	---	---	---
82	---	---	---
83	---	---	---
84	---	---	---
85	---	---	---
86	---	---	---
87	---	---	---
88	---	---	---
89	---	---	---
90	---	---	---
91	---	---	---
92	---	---	---
93	---	---	---
94	---	---	---
95	---	---	---
96	---	---	---
97	---	---	---
98	---	---	---
99	---	---	---
100	---	---	---

```

060 KLFNQLPFLDILLOSFGSGFFPKNS-----SP-----ISVLPQFMGNGMGMI 920
      | : : : | : : | | : | : |
Db 975 NTNPSTSSGGSMSTOTPOSSOSTSPVESTSGATSSSGSGPTLTLSISPSPBSTIG-- 1033

```

```

09      GNGNL-----GNSS--GMIGNSASKPIMPSEBAPQSSAVKRVCAATISAMNRPV   9/6
1033    SSGGSTPVVSHISOGSTDPGSTGTWKPTVSASASSGSGPTMGSTEASTS-----108
Db      ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```

D8  
QY

977 QCGMIRNPAASTIPMRPS-----SQPGHQITLOSQVNNIGPSELEEN-MGGPTQSQQOAPP I03

1088 -GGSSTSPNPNOSTSPTSAGATSSPPSGCT---TTTISTSPSNOSTTGSSDGS----- I13

QY 1031 NOTAPWPESTIPIDQASFSQ-NROPESSPDLLICPHPAAESSEDEGALLDLYALRN 108  
 1138 --TSP---VVSTSGDMTSQSGSTIPSTSGSTVMPQSTGSGSTSGEITSS----- 118  
 Db 1138 --TSP---VVSTSGDMTSQSGSTIPSTSGSTVMPQSTGSGSTSGEITSS----- 118

```

Dh 1185 --CSTOTDPCSTSTSDATSTSTOOSVSMNSDGSMTVMODSTVRGSMGSCSTVMYMGSTEGGS 124
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 1090 FQGLEIDR-ALGIPELVSQSQ-----AVDPEQ 111

```

QY 1117 FSSQDSNIMLEBKAPVFPQYASQAQMAQGSYSPMODPN-----FHTMGGRSXYATLR 116

QY	1170	MQRPGRLPTGLVQNPNQLRLQLHRLQAQQNRQGLMLNQSIVSNVLTLRGG-----	1222
Dh	1302	MDPMTCTSTVTM-----DSTVCGSNGSGCTVTTCGSEASTSC	1333

QY 1224 -----VPTQAPINAQMIAQRREILNQHLRQRQMHQQQVVQRTLMRRGGGLN 127  
|||:::-  
DB 1238 CQVVCPCDSCIDNVDCSDICDNCMAAC CMCQCMTCC 127

QY 1272 MTPSMVAPSCGNPATMSNPRIPQANAQOF-----PPFPNYGISQODDPGFTGA 131

```

OY      1319 TTPOSP LMSPRMAHT-----QSPMMQOSQANDAYQAPSDING----- 135
      : | | : | | : | | : | | : | |

```

OY 1356 WAGGNMGNSM-----FSQOSPHEGQAANTSNNMNINVMATNTNGSSMMNQ 140  
::|: |::: | :| |::|| : : ::| ||

QY	1408	TGQISM	TSVTS	TG	SSMG	PEQV	ND	BALR	GN	LE	PN	Q	L	P	G	M	1450
		:	:	:					:	:	:	:	:	:	:	:	:

## RESULT 6

C;date: 15-Aug-1996 #sequence\_revision 18-Oct-1996 #text\_change 21-Jul-2000

A:Title: The *Drosophila* melanogaster similar bHLH-PAS gene encodes a prote

A; accession: J04831  
A; molecule type: mRNA  
A; Residues: 1-1505 <NAM>



A:Cross-references: GB:043090; MID:g1174073; PIDN:AAC47303.1; PID:g1174074  
 C:Genetics:  
 A:Gene: sima  
 A:Cross-references: FlyBase:FBgn0015542  
 A:Map position: 3  
 F:72-125/Region: helix-loop-helix #status predicted  
 F:171-433/Region: PAS domain #status predicted  
 F:506-635/Region: proline-rich

Query Match 4.28; Score 319; DB 2; Length 1505;  
 Best Local Similarity 18.9%; Pred. No. 6,4e-08;  
 Matches 317; Conservative 221; Mismatches 609; Indels 530; Gaps 73;

QY 13 RAERKRRECPDGLSPKRNTERKNEENKYEELAELEIFANFNDIDNENFPCDCAI 72  
 DB RNNNEKEKE-----KSRDAACRSKSETEIFMELSAALPLKTDVNDL-----DKASV 114  
 QY 73 LKETVQIROIKEEKA---NIDEVQSDVSTGGVYDKALG-----P 117  
 DB 115 MRITAIK-LREMLQVPSLRDCNDIKDIEFAEDQEVKPKLEVTEDMINGABARE 173  
 QY 118 MLEALDGFVVNLBENVVFSNTQYLRNOEELMKNKSVSILVGHTEPVKLLP 177  
 DB 174 LKQTMDFLLVLSHEDIVYSENVEYLGITKIDTLGOIWEYSHQCHAEIKELSL 233  
 QY 178 KSIYNGSGWSEPPRRNSHFNCMLVPLPDESEEGHDNOEAHQ-YETMOCFAVQPK 236  
 DB 234 K-----RELQKVKDEPQNSGVSTHNRDLFVRLKCLTISRGR 271  
 QY 237 SI-----KEGEDLOSCLICVARRVPKERVLP-SSEPTTRDOLQ 277  
 DB 272 SINKSASYKVIHTGHLVYNAKERL-LMAIGRPILPHSNIEILGTSTFLTKHSLD 328  
 QY 278 GKIT-----SLDTS-----TMRAMPCHMEDLVRCIOK 306  
 DB 329 MRFTYVDKMHDLGLYLPKDLDTLSFCOHGADSERLMAFKSVLSKGGGERSR--YR 385  
 QY 307 FNAOHEG-----ESVYAKRHHEVLRQGLAFSQIYR--FSLDGLTLVAQTKSKL 355  
 DB 386 FLGKYGICMILSQATIVYDKLKPSVVCVYVYISNLENKHEIYSLAQOFAASQKQHH 445  
 QY 356 IRSQITNEPOLVLSLMLHREOVNVCNMPDLTGQTMGKPLN----- 396  
 DB 446 QAATEEKEPEKAADPEIIAETKETVNTPIHTSELQAKPLQLESEKYEKTIETKIATI 505  
 QY 397 -PLSSNBPANQ--ALCGNGODMTLSSNINFRINGPKKMGMGMPGSGCGNNHSGM 453  
 DB 506 PVTATSTADQIKOLPESNPKYL-----QAEELIKR-----ENHSGP 545  
 QY 454 QATTPOGSNVALKKNSPQSGPCGNPGQ--PTSMLSPRHRKSPGVAGSPRIIPSOFSFAG 511  
 DB 546 RTIQAQ-----LISGSSSGLRPEKPKRSVTASVLRSP--APPLRP-----PIA 589  
 QY 512 SL--HSPVGV-----CSSTGNSHYTSSNLALQA-----LS 541  
 DB 590 VLCKKTPGLVERNLPTTTATTAALISSNOQLQIAQOTOLQNPQAPADMSKGCFLAD 649  
 QY 542 EGHGV-----IGSSLS-----PDLKMGNLQNS--VYNNMPPLSKMGS 579  
 DB 650 DGRKLTLMKEPPDLNHLNSTNCIDEMTPFSMDMLVGLMGTLCLPEBDIN-----S 701  
 QY 580 LDKSDCKGL-----YGEPEEGTQGAESSCHPGQEKETNPDLNLP-----PAVSERADGQ 629  
 DB 702 LDTTCKTASGCHYQSSSSST--SAPSNTSSSNNSYANBSLPLGTNSTATAPNPBHQ 760  
 QY 630 SRLHDSGGQTKLQLLTTKSDMPERPLASSLDTNKDSTGLSGSGSTHGTSLKEKHKI 689  
 DB 761 QOQHNNQOQOQOQOQOQOQOQOQHNPQHNDNSNSSNIDPLENY--REESNDTSCSQH-- 812  
 QY 690 LHRLLDSSSPVDLAKLLEATGKDLQSESSSTAPGSEVTTKQRPVPRKKENMLLYLL 749  
 DB 813 LHSPTSKSPED--SLSPSLCSPNSLTQEDDFS--FEAFAMRAPYIP-----I 857

QY 750 DKDDTKDIGLEPITPKLERLDSKTPASNTKLIAMKTEKEKMSFEPDQPGSELNLEET 809  
 DB 858 DDD-----MPLTEFDLMMCPPEDLQTVWR-----ETDALQO 891  
 QY 810 LBDLQNSQLPQLFPDTRPGAPAGVDKQALINDLMQLTANSVYTVYGAQKTAALRSOST 869  
 DB 892 LQQLQOQOQHQQY-----ASN--TGYQOQOQOQOQOQOQOQOQ 923  
 QY 870 FNNPRPGQLGRLLPNQNLPLDITLQSPGAPPPPIRNSSPYVIRPQGMKMGKMG 929  
 DB 924 FSNM-----LCSSPA-----STVSSLSPSPVOQHQ----- 949  
 QY 930 GNIGNSGTMGINSASRPMPGEMAPQSSAVVTCATTSAMNRPVGGMIRNPAASIP 989  
 DB 950 ---QQAAYVTSDSSELAALLCGSGNGTJSLILAGSGVTVEECNERLO----- 994  
 QY 990 MRSSOPGQROTLOSQVMNIGPSELEKNNMGPPQYS-QQAAPPQOTAPWESILPPIQAS 1048  
 DB 995 ---QHQQOQOQOQSGNEFRFQQLQOELOEBOORQOQOQOQOQOQOQLSLTECK 1051  
 QY 1049 ASQNRQPFQSSPDDLCPHPAES-----PSDEGALLDQLYLALRNPGLFEIDRALGIP- 1103  
 DB 1052 KEKTYDQMGSS---LC-HPMEDAFENDYSKDSANLDCMDLQOYVDTPEVSPNPAASP 1106  
 QY 1104 -----ELVSOQAVDPEQFSSQDSNIML-----EOKAPVEPQOYASQQAQGS 1147  
 DB 1107 PKVYSAILQLQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQ 1162  
 QY 1148 YSPKQDPNFTMGQRPSYALRMOPRRGLRPTGLVQNP--NQLRLQDLRLQAOQNNQ 1204  
 DB 1163 IKLINGASLAPVNTK--ATIR-----LVESKPPTTQSRM-----AKVNLV 1201  
 QY 1205 PLMNQISNVSNVNLTLRPV-----QAPRNAQML-----AQOR----- 1240  
 DB 1202 PQOQOQHNKRNHLNSATGAGPVPESKRLKSTGLDQVQSPOLLQOLIGKBPAAQOQQAAR 1261  
 QY 1241 -----EILNQHLRQRMHQOQOQVOQRTLMRGQ-----LNMTPSNVAD--SG 1281  
 DB 1262 AGSEBRLQSAESKQKQOQOQNSVSLKLLVSGRDDDSSEAMIIDEDNSLVQPLGKYG 1321  
 QY 1282 MP-----ATMS-----NRIPOANAQOPPPPNVG-----ISOQDP 1313  
 DB 1322 LPLCHTSTSSVLDYHNHNLISGTN--FOLSPVEGSDSSGDEBTSVSLDSVPP 1378  
 QY 1314 GFTGATP-----POSPLM-----SP--RMAHTQSPMQOQANPAYQAPDINGMA- 1357  
 DB 1379 GLTACDTDASDSGIDENSLMDGASGSPKRRLSTSTNSTQOASAPRALDVERTPVQKSV 1438  
 QY 1358 ---QGMGNSMFPQOQSPPHFGQOANTSMYNNMN--INVSMAVTGMS-SMNO 1407  
 DB 1439 EELEEGCGSGNNAPSRKTSISFLDSSNPDLHTFAMMDLVDDYIMGCGFEGSGNOL 1495

RESULT 7  
 T03455  
 ALR protein - human  
 C:Species: Homo sapiens (man)  
 C>Date: 24-Mar-1999 #sequence\_revision 24-Mar-1999 #text\_change 21-Jul-2000  
 C:Accession: T03455  
 R:Prasad, R.; Zhadanov, A.B.; Sedkov, Y.; Bullrich, F.; Druck, T.; Rallapalli, R.; Ya  
 Oncogene 15, 549-560, 1997  
 A:Title: Structure and expression pattern of human ALR, a novel gene with strong homo  
 A:Reference number: 214954; M0ID:97388474  
 A:Accession: T03455  
 A>Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residue: 1-4957 <P>  
 A:Cross-references: EMBL:AF010404; MID:92358286; PIDN:AAC51735.1; PID:92358287  
 C:Genetics:  
 A:Gene: ALR  
 A:Map position: 12  
 C:Superfamily: human ALR protein



[illegible][illegible]

RESULT 10  
H96747  
unknown protein T10D10.14 [imported] - Arabidopsis thaliana  
C/Species: Arabidopsis thaliana (mouse-ear cress)  
C/Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
C/Accession: H96747  
R/Theologis, A.; Ecker, J.R.; Palm, C.J.; Federpspiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chih, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Cressy, T.H.; Dewar, K.;  
ansen, N.F.; Hughes, B.; Hultzer, L.  
Nature 408, 816-820, 2000  
:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.

Query Match	3.9%;	Score 297.5;	DB 2;	Length 1088;
Best Local Similarity	21.3%;	Pred. No. 4.7e-07;		
Matches 259;	Conservative 146;	Mismatches 444;	Indels 365;	Gaps 65;

OY	377	ONVCVNM-PLDLE---GOTMGPELMDISNSPAHOL-----CSGNPODDMTSSINIFPLNG	429
Dd	15	KKVCIDRLPESSEENGNIPLHILIMQOTRNNQATONLGSTNMLAGLRQPLQDAPNSSLALVP	74
OY	430	PREQ--MOMPEGRPGSGSGGNHVS-----GMOATTPQGSNYALKNPSPOSSPGMNG	480
Dd	75	POQQRMMIGISTRTNTPQOGNSVSVSGASHPGIGLDALPLPGSD-----SNMFG	121
OY	481	QPTSMLSERH-----RMS--PGVAGSPRIIPPSQFSPAQSLHSFVYCSSTGNSHTYNSL	534
Dd	122	--TSFHRKRESQEGOMSMGMLKRTTRV--SHMGPGGVPOQOLQO--RMDGLHGSDTNMKN	176
OY	535	NALQALSGHGVSLGSSLASPLDKMG-----NLQNSPYMMNPPRLSKMSL--D	581
Dd	177	TLLO-----HODMLGNITQYPTNTSIQRFSPHQMEGYMNOEGGPMPF--PASQOGMKYT	228
OY	582	SKDFGLYGEPSSEGTGTQOASSCH---PGBOKEITNDPNLPPAVSSERA-----	626
Dd	229	SKE-----EPFE--TGKIDGTRNNIPGVGSAND--LDPRIGSRMHPNMFIRSNPQT	278
OY	627	-----DGOSRLHDSKQGTKLLDILLTKSDOME-----PSPLASSLSDTNKOSTG-----	670
Dd	279	SWNVPGOOIEKEPKKEQFSRRISQASOPLRSLAGBPQPSLSSKGFECSGSMGTHYGA	338
OY	671	-----SLPSSGTHGTSLKEKHKLHLRLDQSSPYDLATLTAEAGKDLQSSSS	720
Dd	339	AAAOCKDAVYISIPIGATQSVSSANEAMQORHQ-----AOAMAKRTNRTSLPKYOV	390
OY	721	STARGEVYTK--OEPPSPKKENALLRYLLDKDKIDKIDGLPEIRPK-----LERLDS	771
Dd	391	ISTYGPSVSVNTISVPMANS-----PSVGPQLGSHALLDRSK	430
OY	772	KTDPASTNTKLIAKTEKEEMSFEP---GDQPGSE-LDNL--EELIJD-----	812
Dd	431	IERVAAROYLNCNKKHKKVDEFSRRPRVAKOPLVLCISLNSNEEVEKDEDEALSKSIFGS	490
OY	813	-----LQNSOL-----POLPDPTR-----PAPA---GSYDQ	837
Dd	491	MNTKTRVIFHGOMERAMQDSSIFSSSVSFLPRNFTRLVMSEKAVDGTVMATQGDVE-	549
OY	838	AIINDLMQLTAEANSPLY--PYGAOKTALRISQSTFNNRPGQGLRLLPNQNLPLDITLQSP	896
Dd	550	--GDVFO--AEDPLLALPYTMAREGYMEHIMAKPNRGDGT-----PIS---SHP	593
OY	897	TGAPFP---PIRNSPYSTIIPQGMAGNQMIGNUGNLGNSST--GMIGNSASRPTMPSG	952
Dd	594	NSAGGYRGYSANDMOQYGDVAVAGQASGASKRGNTGNTPNNTSTNILLANRMPV--PTN	651
OY	953	EMAPQSSAVRVCTAATTSAMNRPVQGMIRNPAASIPMPPSSQPGOROTLOSQYMNIGPS	1012
Dd	652	SQALQMSQ-----GLL--SGVSPMPOQOOLDPQOQSALLS-----S	684
OY	1013	ELENMNGPOYSQOAPPNQOTAPWESILPIQDASFASQNRNRPFGSSPDDLLC-----	1065











OY 976 VOGGMRNPASIPMRSPQPGOROTLQSOVMNIGPSELEMMNGCPQYSSQOAPRNOTAP 1035  
DB 705 -----RITPQSLNPG-----QMSMAQPRPVPRQPTLQ---- 734  
OY 1036 WPESILPIDOASASQNRQPGSSPDDLCPHPAESPSDEGALLDQYLALNPFGL 1095  
DB 735 -----HNGQLAGCALNPP--MGYGRMQQPSNQGOFLLPQ----- 767  
OY 1096 IDRALGIPELVYSQAVDPBQFSSQSDSNITLBEKARVFPQYASQOMAGS---YSPK 1152  
DB 768 -----TQFPGQGMVNTNIPRLAPSSGAPVSOAMSSSCPVNSPI 808  
OY 1153 DPNFHTMGORPSYATLRMOPRGLRPTGLVQONPNOLRLQLQRLAQOQRQPLMNQSN 1212  
DB 809 PP-----GSGSHIHCROPQLRH-----QNSPSP----- 834  
OY 1213 VSNVNLTRPVPPTQAPRNAQMLAQRRELLNQLRQMHQOQOYQORTLMRGGLM 1272  
DB 835 -----VPSRTP-----TPH-----HTPESI-----GAQ 853  
OY 1273 TPSSVAPSGMPTMSPRIIPQANA-----QQPFPNYSIQQPDGFTGATTQSPFLMS 1327  
DB 854 PPAFTTIPADVPYPPAMPQPGQSLHPRPQRTPTTQDLPQOQVQBSLPAAPADQPQQ 913  
OY 1328 PRMAHT-----QSPMQOQSQANPAYQAPSDING 1355  
DB 914 PRSQGSTAASVPTPNAPLPRQPATPLSQPAVSI 949

RESULT 16  
T20532  
hypothetical protein F07A11.6b - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 29-Oct-1999  
C:Accession: T20532, T27777  
R:Palmer, S.  
submitted to the EMBL Data Library, October 1995  
A:Reference number: Z19287  
A:Accession: T20532  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-2722 <M12>  
A:Cross-references: EMBL:266511; PIDN:CAB54211.1; GSPDB:GN00020; CESP:F07A11.6b  
A:Experimental source: clone F07A11  
R:Galadsky, S.  
submitted to the EMBL Data Library, March 1996  
A:Reference number: Z20417  
A:Accession: T27777  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-2722 <M12>  
A:Cross-references: EMBL:269904; PIDN:CAB54502.1; GSPDB:GN00020; CESP:F07A11.6b  
A:Experimental source: clone ZK20  
C:Genetics:  
A:Gene: CESP:F07A11.6b  
A:Map position: 2  
A:Introns: 36/2; 92/2; 182/2; 272/3; 344/2; 426/2; 490/1; 541/2; 700/3; 770/3; 1286/3; 1

Query Match 3.6%; Score 275; DB 2; Length 2722;  
Best Local Similarity 19.6%; Pred. No. 2.2e-05;  
Matches 333; Conservative 220; Mismatches 694; Indels 450; Gaps 76;

OY 6 ENTSDSRATETRRKREPDGLSPKRNTEKRNRECNKYTEELAEILFANFNIDINF 65  
DB 1028 EARRERKRETKERNRKRKRERAKRLEDEROEKREK--KERDEK----- 1072  
OY 66 KPDKCAILKETVQIQRKQEKRAAANIDEVQSDVSSTGCGVTD-KDALGPMLEALD 124  
DB 1073 KRKEKVRKAKEKELKKKKHKG-----DSSESDDSDNDELDDIVRKTKETKQGEKD 1127  
OY 125 GEFVVVNLGCVVGVSENVTOYLRYNO-----EELMKNKSVSILHVGHTFERVKNLLPK 178

DB 1128 HQALLLSKGIIT---ENLKSRRSRDKRAHDSFEKMQKS-----QQRVLIES 1173  
OY 179 SIYNGG-----SMGGE-----PPRNSHTFNCRMILYKLPDSE----- 211  
DB 1174 SDDEGKDKDGKNGSSNGEESDSEKADLPPRPAPSPISSESIDOKLAKVKEKELTTSSD 1233  
OY 212 -EEGHNOEAHQKYEFTMOCFAVSOQSIKEEGEDLOSCLICVARRPMPKERVLPSSSEF 270  
DB 1234 DEHDNAGELHQRLREDENRKRQKSLTAYSSDEG-----ERKNVP----- 1276  
OY 271 TTRDLOGKITSLDTSTMRAMPGW---EDLVRCIQKFNHQ-HEGESVSTAKRHHEV 326  
DB 1277 -----KMRRDGEDDAAKHPGWSAKDDQKQRKRKLEHRRSDESKNAKRPFRDI 1328  
OY 327 LRQGLAFSOLYRSLSDGTLYAQTKSLRQTTNE-----QQLVLSLMLHREDN 378  
DB 1329 PHEDVSDDE---ETEDGSKSRROSTSTISNTAKEREKSGTPLYRV----- 1374  
OY 379 VCVANPDLTGOTMGR-----LNPISNSPA-----HQALCSGNPQ-DMTLSNI 423  
DB 1375 -----PEPTPTPLSKILSPKILSPKTSSTSKRSSIDHEMLIS-PRQRRTTSSIS 1427  
OY 424 NFPINGPKQMGMPKGRFGSGGMHVSQMATTPOGSNTALKMNSPSSQSGMNGOPT 483  
DB 1428 TATTSKHEALSIPEKPL--SPVTAKSSVSSIDPSIRDEFMNS-AADSPMSTGRP- 1483  
OY 484 SMLSPHRMSPGVAGSBRIPPSQFSPAGLSHPVGCSSTGNSHTNS----- 533  
DB 1484 -WVLTAKAMK--AFNSTPKKKNNSSGQHDSSSGSSSDSGSTSDSDSDDEVPKO 1539  
OY 534 --LNLQALSEGHVSLSSLASPLDKMGNLQNSPVNNPPLSKMGLSDSKDCEGLYG 590  
DB 1540 TEPTVTSIPVVASDNGSPENVVETPSI---VQGTREPREPTISQSS-----BS 1586  
OY 591 EPBEGTTQAESSCHGEQKETNDPULPRAVSEBRD-----QGSRLHDSKGQTKLQDLT 646  
DB 1587 EP-EAVPECEASVER--QMETSQ-NVEP-VSEHDSHGHGSEVAVESQDQPL----- 1636  
OY 647 TKSDMEPSPLASISLDT-----NKDST-GSLPQSGSTHGTSLEKIKILHRL 694  
DB 1637 --EHQEKELKNIIDVAEHNHEBOYQDEDESVESIRAPSEPPRYQAQKSAHTLI 1694  
OY 695 QDSSSPVDLAKILAEATGKDLQSESSSTAGSEVITKQEVSPKKEMALLRYLDDKDT 754  
DB 1695 SDQETQAVQSIDPEEADFPQ-----YPDFISTNEKEVSGKPDHNIKTEPLNNGHT 1749  
OY 755 KDIGLEPIT--PKLERLSDKTPASATKILAKTE-----KEEM 791  
DB 1750 DILFSPSSAHASEKOSTKSEDMEDSESLVMEKEVPMQVIAQEVHVPSEPSMEEV 1809  
OY 792 SFEPGDQPGSELNLEIILDDLQNSQLPOLFPPTRGAPASVDKQAIIN-DLMQLTAE 850  
DB 1810 KLETSVPKPEPIKME--SPEQTPRLDISNNESODTGAANNHLHENHDAVQ----- 1861  
OY 851 SPVTPVGAQKRTALRISQSTFNNRPQOLGRLLPN--ONLPDLTILQSPGAGPFPPIRNS 908  
DB 1862 --TPQLQASQNH--QVAPSPSPA-----VAPDSQONGVILVSQS----- 1899  
OY 909 SPYSVLPQPGMNGMIGOGNLGNSSTGMIGNSASRPTMPGSENAPOSSAVRYCAAT 968  
DB 1900 -----QPSPMSSQOSDMAQ-NLILSSKIDINDLAKILKRNPA---LAQATRGDSGT 1947  
OY 969 TSAMNRPVQCG-----MIRNPAASIPMRSPQPGOROTLQSOVMNIGPSEL-ENMMG 1020  
DB 1948 FOHLLHAQNGMGMTRPMLQLKAFFAQOQENEAQMMQAKKOQTIINKDRKEDERYK 2007  
OY 1021 PYSQOQAPRNO-----TAPRESILPIDOASASQNRQ--PRGSSPDDL 1063  
DB 2008 RMVEENRKYVEDRREKORKEERORLAATAATATGATQAAABALAKOKQVRHNGOHV 2067  
OY 1064 LCPHPAESPSDEGALLDQYLALRNFQGLEETDRALGIPELVYSQAVDPBQFSSQSDSN 1123

Db 2068 SMPTPEARS-----LYEOPGGLSSYINRDSIGATNGVLTLPQOS-IQRPSTASTSSN 2119  
QY 1124 IMLEOKAPVFPQYASOAOAGSYSPMODNPHFMGORSYATLRBMQPRGLRPTGLVQ 1183  
Db 2120 ---PPKAPLQD-----SASVQNTIDPAIEIEIRV--QRMFTKPLKMSAEAAIYMAVAS 2169  
QY 1184 NQPN-----QLRLQLOHRLQAOQNRQ-----PLMN--OISVSNVNL 1218  
Db 2170 SDPWPATSTVDLAAMQ-QQAOAAQAQAQVVPVYTTASTPRLSNLETLLSTASTLANL 2228  
QY 1219 -----TLRPVPTQAPLINAQMLAORRELLNOHLROR--QMHQOQOVQ 1259  
Db 2229 ATGALNPLSMALTSLSQSSPYOGIARVLLTMNQMLATQTSSELLATMNOQETLM 2288  
QY 1260 QRTLMNGOGL-----NMTSPVAVPSC--MPATMSNRLPQANAQOF----- 1299  
Db 2289 ---ALLAARNGLPFAMPQOQOQOPMPAOGGFAIPVLPMSIKRNAKDDLSSGVYSDRKS 2346  
QY 1300 -PPPPNYGISQDPDPTGATTPQSPPL--MSPRMAHTQSPMAQOQOANPAYQAFSDINGW 1356  
Db 2347 CPLHAMIQGGQQPP-----PQPMQAVAPAPRSPSPRKGMFEN--LPEPMK-- 2393  
QY 1357 AQCMGNGSMF-----SQQSPHFGQOANTSMYNNMNIIVSMAT 1396  
Db 2394 -----EKNEMRKEILRLDITLLELGADEEDQKPDLKQ-----IPTSEDTDDSKAD 2443  
QY 1397 NTGG-----MSSMNMTGOISMTSVTSTGSLSMGPBQVNDPALRGNLFPN 1445  
Db 2444 SMGEGSAFRRILSRSTSMGNSGSPSASGTSTSSISSGPD---SPLEGE---PL 2497  
QY 1446 QLPQMDMIKQGDYTRK 1462  
Db 2498 NSEFMDLTFEVAQKHRE 2514  
RESULT 17  
B86292  
F7H2.12 protein - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C>Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 02-Nov-2001  
C:Accession: B86292  
R:Theologian, A.; Eckert, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;  
ansen, N.F.; Hughes, B.; Hutzler, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maitli, R.; Marziani,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719  
A:Accession: B86292  
A:Status: Preliminary  
A:Molecule type: DNA  
A:Residues: 1-1366 <STO>  
A:Cross-references: GB:AE005172; NID:98927657; PIDN:AAF82148.1; GSPDB:GN00141  
C:Genetics:  
A:Map position: 1

Query Match 3 64: Score 273.5; DB 2: Length 1366;  
Best Local Similarity 18.58; Pred. No. 1e-05;  
Matches 252; Conservative 186; Mismatches 459; Indels 461; Gaps 63;

QY 261 RPYLPSSESTTRQDLOGKITSLTSTMRAMKGMEDVLRICQKFAHQEGESVSAYAK 320  
Db 7 RPLSPNCE-----PAMDTGWRQQLRP---DSRQIKYKNI----- 38  
QY 321 RHHHEVLRQGLAS-----QYRFSLS-DGLVAAQTKSLINSQTTN----- 362  
Db 39 ---METLKHLPFSGPREGINELRIARAFEEKIFSGALNORFQWTPPOHGKEITFGICK 95

QY 363 -EPOLY-----ISLHLHRE---QNVCMNPDLTGOTMGKPLNPISNSPAHOA 407  
Db 96 AKPOVYGEIHTDYLRKISMKMLMETKSQAAGSSAAIPRANNTSTDSITPNO--GQL 153  
QY 408 LCGNPGQDMLSSNINFPINGKEQMGMPGREGSGGMNHNVSQMATTPQGSYALKM 467  
Db 154 L---PG---SLSTNOS---QAQPLLSQTMQN-----WTASQMTGST-----ALPS 190  
QY 468 NSPQSSSGMNPQGTSMLSRHRHSPVASSPRIPRQFSPAGSLHSPVGVCSGTGSH 527  
Db 191 SMPVYS---ITNNNTSVNQN-----ANQVAVAMLDSSGQH 227  
QY 528 SYTNSSLAA--LQALSESHVS-----LGSLSPLDKMGNLQNSPVNMP 572  
Db 228 GLSSMFGPQROMLGRHAMSQQQOQPPYLXQOLOOLKQNFQOSQNVN--PMSLPS 286  
QY 573 PLSKKGLSDKDCGLYGPSEGTGQAESSCHPEQKEITNDPLPPAVSERRADGOSRL 632  
Db 287 HIOQ-----QQQNVLDPRNQLHSSQOPGVPTS----- 312  
QY 633 HDSKQGTLLDQLTTKSKMEMPSPPLASSLSDTKNDSTGSLPGSGSTHGTSLKEKHILHR 692  
Db 313 -----ATQPSVNSAPLQ--LHTNODSSPOLSSQOYTQOS-----M 346  
QY 693 LLDSSSPVDLAKLTAEATGKDLSESSSTAPGSEVITKQBPVSP--KKENALLRYLLDK 751  
Db 347 LRQHOSSMLRHPQSOQASG--HQOQSS-----LPQOSIFLQOQPPOLNR----- 391  
QY 752 DDTKDGLPEITPKLERLDSKTPDASNTKLAMKTEKEMSFERQDQGSSELDLEETLD 811  
Db 392 -----QQANSSGI---QKQKMG-----QHYVG 412  
QY 812 DLQNSQLPQLPDRPRCAPACSVDKAIIINDLMQLTAENSV--PRVCAQKALRISOSTF 870  
Db 413 DMQOOHQRL-----LQONNNVMNIQOQOQOQOPLQOPOQOQK----- 450  
QY 871 NNPRPGQLRLPNQNLPLDTLQSPGAPPEPPIRNSPVSVPYIPQPMGNOGICNOG 930  
Db 451 ---QOPPAQOQLMSQON--SLQATHQPLG-----TOSNVAQLOP---QOQMLNSQ- 494  
QY 931 NLGNSSTMGNSASRPTMGEMAPQSSAVRYTCATTSAMNRPVQGMIRNPAISPM 990  
Db 495 -VGNSSLQNNQSHVHMLSQP-----TWGLQRTHQAG--HGLYSSQGG 533  
QY 991 RPSSQPGQQR---TLOQSVYNTIG--PSELEMMNGGQYISQOQAPRQGTAPWPESTLP- 1042  
Db 534 QSONQPSQOQMMPLQSHHQDLGLQOQPNLLQOQV---QQRRLASGQVGT---SLLRP 585  
QY 1043 ---IDQASFASQNRQPFSSSPDDLCPHRAESPSDEGALLDOLYLALRN-----PQGLE 1094  
Db 586 QNVVDQQRQLYQOQRTLEMPSSSLDSTAQTES--ANGQDQOEYVQKIKSMKETYLPLDLN 644  
QY 1095 EIDRAL-----GIPELVSQSOAVDPEQF-----SSQDSNIM--LEOKAPVF- 1133  
Db 645 EITQRYVAAKLOQDSDMPQQRSDQLKRLQFKTLMERMIOFGLSVYSKSNIMRLKQVAYTE 704  
QY 1134 -----PQOYASOQAQAGSYSPMODNPHFMGORS--PSYATLRMQ----- 1171  
Db 705 KQIIGFLNMHRPRKRVQOQGLRPOQOMQOPQOSQOTVQDOSHDQNTQPNQMSKMGAGAP 764  
QY 1172 -----PRPG-----LRTGLVQONQPNQDLRLQLOHRL--QAQONR 1203  
Db 765 RAQOSSMTNMQSNVLSRPGVSAPQONIPSSIRPASSLESQGNLNNQOAVAGSMQONT 824  
QY 1204 QPLNQISNVSNVNLTLRPGVPTQAPLINAQMLA-----ROREIILNOH 1246  
Db 825 SOLVNNSSASASQSGISTLQSNVNVNQPOLSSSLQNHQKLQOQDQOMQLKQRPQOQOMQOQ 884  
QY 1247 LRQNMHQOQOQOQOTLLMRQGLAMTPRSMVAVRSGMPAT--MSNPRIRPQANAQOFPPPPN 1304  
Db 885 LQARQOQOQOQOQOLARQOAAQLOQNMNDLTSRQGMNVSRRMFOHNSMOGQARVPL--- 941  
QY 1305 YGISQOPDPRTGATTPQSPMLSPRMAHTQSPMAQOQOANPAYQAFSDINGWAGNGGN 1364

Db 942 ----DQLKPG-----AVSSPOLLOGASPMQSG-HLSPVDOKNTNKMGTPLQOPAN 987  
QY 1365 SMFSQOSPHFGQOANTSMYNNMNNVSMATNTGMSMN-----OMTGOISMTSTV 1417  
Db 988 SFFVYVSP-----SSTPLAPSPMOYD-----SEKROSSLSMGNARQOATGMQGVQSL 1037  
QY 1418 SVSTGSLSS-----MGPEQ--VNDPALRGCNLPNPOLP 1448  
Db 1038 AIGTPOISASPLLOEFTSPDGNILNSTITSGKSPATELP 1077

## RESULT 18

JE0270

A:Ant-like PAS protein, Arnt3 - mouse

C:Species: Mus musculus (house mouse)

C:Date: 05-Feb-1999 #sequence\_revision 05-Feb-1999 #text\_change 21-Jul-2000

C:Accession: JE0270

R:Takahata, S.; Sogawa, K.; Kobayashi, A.; Ena, M.; Mimura, T.; Ozaki, N.; Fujii-Kuriyama

Biochem. Biophys. Res. Commun. 248, 789-794, 1998

A:Title: Transcriptionally active heterodimer formation of an Arnt-like PAS protein, Arnt3

A:Reference number: JE0270; MUID:98369629

A:Accession: JE0270

A:Molecule type: mRNA

A:Residues: 1-626 &lt;TAK&gt;

A:Cross-references: DDBJ:AB014494; NID:g3402482; PIDN:BAA32208.1; PID:g3402483

Query Match 3.6%; Score 272; DB 2; Length 626;

Best Local Similarity 24.1%; Pred. No. 4e-06; Mismatches 229; Indels 130; Gaps 28;

Matches 141; Conservative 84; Mismatches 229; Indels 130; Gaps 28;

QY 35 EKRNRQEKYIEELAEILFANFNIDNENFKPKCAILKEVYKQIRQIK-EOEKAAAN 93  
Db 81 EKRRRQKMSFIDELASLV-----PTCNAMSRKIDKLTGLRMVQNHKTLRGANPTTEAN 136  
QY 94 IDEVOKSDVSTGGQVIDKDALGPMLEALDGFEEVYVNE-GNVFVSENVTOYLRYNOE 152  
Db 137 YKPTFLSD-----DELKHLIRAADGFLFVVGCDRGKILFVSSEVKILNYSON 185  
QY 153 ELMNKSIVSILHYGDHTEFEVKNLIPKS-----IYNGSGWSEPPR 192  
Db 186 DLIGSLFLYLRHKDIYAK-VKEQLSSSDTAPRERLIDAKTGLFVKDTIDTPGSRLLCSGAR 244  
QY 193 RNSHTENCMLV-KPLPDESEEGHNOEAHQYETMQCFVAVSOPKSIR----- 239  
Db 245 R--SFGCMKCKRSPVYKEDKDPASTGSKKADRSFCTHSTGYLKMSPPTKMGLED 301  
QY 240 ----EEGEDLOSLCIVAR----RVPMKER-PVLPSESEFTTRQDLOGKITSLD-TSTMK 289  
Db 302 NEPDNMGCMV-SCLVAIGRLSHMVPQANGELRVKSMEVSRHAIDGKFVVDORATAI 360  
QY 290 AAMKPMEDLVRCIOKFNPAHQHGESVSYAKRHNEVL--ROGLATSOIYRFSLSGTLY 347  
Db 361 LAYLP-QELLGTSCEYFHD----DIGHLAECHEOYLOTRKITY-TNCKFKIKDGSFL 414  
QY 348 AAOQSKILRSQTTNEPOLVISLMLHREQNCVMPDILGTQMGKPLNPISNSPAHOA 407  
Db 415 TLRSRFSEFMNPWKREYIVS-----TNVVLANVLEG--GDPTFP-QLTAPPH-- 461  
QY 408 LSGNGQDMTLSSNINPFLNGKEDMGMPMGFGSGGGMNHSV---MQATTPOGSNTVA 464  
Db 462 ----SMOSMLPFGSGGGRKTHPTVPGIPGGRAGAGKGRMTAEIMEIHRIRS--- 512  
QY 465 LKMNPSOSSPGMNPQOPMSLSPRHMSPG-----VAGSPRIIPSOQSPAGSLHSPVG 519  
Db 513 -----SPSSCGSSPLNITS-TPPDASSPBGCKIILNGTPTDIPSTGLLPGQAQETP--- 562  
QY 520 CSTGNSHSYTNLSNALQALSEGHCYSL-----GSSLASPD 556  
Db 563 -----GYPYSDS-----SIGENPHIGIDMINDNGGSSSPSND 596

RESULT 19  
E88320  
protein F07A11.6 [Imported] - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 10-May-2001  
C:Accession: E88320  
R:Anonymous, The C. elegans Sequencing Consortium.  
Science 282, 2012-2018, 1998  
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating bio  
A:Reference number: A75000; MUID:99069613; PMID:9851916  
A:Note: see websites genome.wustl.edu/gsc/C\_elegans/ and www.sanger.ac.uk/Projects/C-  
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999;  
A:Accession: E88320  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-2738 <STO>  
A:Cross-references: GB:chr\_II; PIDN:CA93781.1; PID:g3881547; GSPDB:GN00020; CESP:F07  
A:Map position: 2

Query Match 3.5%; Score 270.5; DB 2; Length 2738;

Best Local Similarity 19.0%; Pred. No. 3.8e-05; Mismatches 619; Indels 471; Gaps 69;

Matches 308; Conservative 225; Mismatches 619; Indels 471; Gaps 69;

QY 6 EMTSDPSRAETRRKRCEDPOLGSPKRNTERKRNREQENKYEELAEILFANFNIDNENF 65  
Db 1028 EAREKREKRETERKREKREMERAKLEDEDEREKREK--KERDR----- 1072  
QY 66 KPDCAILKEVYKQIRQIK-EOEKAAANIDEVOKSDVSTGGQVID-KALCPMLLEAID 124  
Db 1073 KKEKEKVRKAKREKELKKKKHKG-----DSSDESDSDNDELDDIVRKSTEMQOEEND 1127  
QY 125 GFEFVYVNLGNNVVESENTQYLRYNO-----EELMNKSIVSILHYGDHTEFEVKNLIPK 178  
Db 1128 HOLATLSLKGIT--ENLKSRRSDKRAHDSFEKKQOKS-----QORRYLVES 1173  
QY 179 STVNGG-----SWSGE-----PPRNSHTFNCMLVKPLPDE----- 211  
Db 1174 SDEGSGKDGDKGSSNGESDSEKADLPAPAPSLSESADQRLKYLKREKGETLTSSD 1233  
QY 212 -EEGHDNOEAHQYETMQCFVAVSOPKSTKEEGEDLOSLCIVARRPMKERYLPSSSEF 270  
Db 1234 DEHDNDAGSIHQRLTEDEENRKRKOKSLTAVSDEOG-----ERKNVP----- 1276  
QY 271 TTRQDLOGKITSLDPTSTMAAMKPGW--EDLVRCIOKFNPAHQHGESVSYAKRHNEVL 326  
Db 1277 -----KMRRDSEEDAAAHKPGMSAKDDQKQRKRLKLRSSDESKNAKRFDDI 1328  
QY 327 LROGLAFSQTIRFSLSDGTPLVAAQTKSLIRSQTTNEPOLVISLMLHREQNCVMPDIL 386  
Db 1329 PHEDVDEE--ETEDGSRSRROQSTSTISNVTAK-----KREKSGKTPLR 1373  
QY 387 TQGTMGKPLNPISNSPAPQALCSGPNQDMTLSSINPFLNGKEDMGMPMGFGREGSGG 446  
Db 1374 VPEPTGTPILSPKILSPKH--LSPKSTSTSTKRSIS--DHENLLSPROR----- 1419  
QY 447 MNHVSQMAQTTPQGSNY-ALKMNSPQSSPGMNPQOPTMLSPRHMSPGVAGSPRIIPS 505  
Db 1420 -NRTTSTTATITTSKHEKLSIPEKLPSPYVAKSSVSSIDDPSTIR-----D 1465  
QY 506 QPSPAGSLSPVGVCSSTGNSHSYTNLSNALQALSEGHCYSLGSSLASPDLMKMNIONS 565  
Db 1466 EFSMNSAASPMP--STGRPVYLTKA--AKKAFNTPPKKYVSLIIDCYMLGMAMN 1519  
QY 566 PVNMPNPPLSKMGSLDSKCCFGLYGEPSGTTGQAESSCHPGEOKETNDNPLPRAVSSR 625  
Db 1520 -----SSGQHDSSSSGSSSDSSSDGSSSTSDSS--DDEVPKQTEFVTSIPVVASQN 1569  
QY 626 ADOGSRLLDSKQOTKLQLLTTRKQDMEPSPLASSLSLNNKSTGSLPSG-----ST 678  
Db 1570 GSPENVVETP-----STVSQTPREPEPTTISQSSSESEPAVPCPRAVSPEOMETSO 1623

```

Oy 679 HGSLKEKHKILHR-----LLQDSSSPVD-----LAKLTAE-----ATGKDL5 716
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1624 NVEPVEEHEHDEHGESEVAEVSQOQPLEHGEKELEENKLTIDVAAEHHEHGOVGDDE5 1683
Oy 717 QESSSTAAGSEFTIKQEVSPKKNALLRYLLDKDDTKDGLPITTPKLEHLSKTDPA 776
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1664 VESSSTPAGSDE---PDVYTAQEKSA---HTLISDOETDQAVOSIFD---EEEADEF 1731
Oy 777 SNTKLIANKTEKEMSFPPGDDQSELDNLEILDDLONSOLPQFPPTRPAGASVDK 836
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1732 POYPPFGISTNEKVS-----GKDPNIRK-TEPLNNGHTDLF---SPSSSAHASEK 1780
Oy 837 QAI-INDLMOLTAE-----NSPYTPVGAOKTALRISOSTENNPPRGOLGRLLPQNLPL 889
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1781 QSTKSEDDMEEDSELVMEKEVPMQVIAQEVH-----VPSSESPM 1821
Oy 880 DITLQSPGAGPFP---PIRNSSPYSVIPQGMGNQGMIGNQGLNNGSSTGMIGNSASR 946
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1822 EEEVKLET--SPVPKEEPIKKEESPEQTPPTDLISN-----NESQDTPGAVNNHLE 1871
Oy 947 PTMPSGEVAPOSSAVRYCAFTTSMNRPYOGMIRNPASTPMRPSQPOGROLQSOY 1006
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1872 -----NHDVQTPDQ-----LQPAS-----QHOY 1890
Oy 1007 MNIGPSEL---EMNMGPOYSQOQAPRQOTAPWPEILPIDQASFAQNRQPFSSPDL 1063
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1891 AQPSPRAVAPDSQONGVLYVSQOSQPS-----PMS-----SQOSMAONLLISSDIDNL 1941
Oy 1064 ---LCPHAAEPPSDEG---ALLDQLYALRNFQGLEIBRALGIPELVSOQAVPBOF 1117
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1942 AAKLHKNEALAOATRGCSCGIFPHLLHAQ-----GNGQNMTPP-- 1981
Oy 1118 SSQDSNIMLEOKAVYFPOQVYASQAOAGSYSPMODPFRHMGORPSTATLRMQRPELR 1177
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1982 -----MLQKAAPFAQOOEENAOQMAQKQOTINKDRIKEOE-----R 2021
Oy 1178 PTGLVONOPNQLRLQLOHRLQAOQNRQPIQMNQISNVNLTLRGVPQAPINQMLAQ 1237
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2022 VKRYAEERKEVEEDRRKQKKEERORL-----AAATAATMATQ 2062
Oy 1238 RQRETLNOHLRQORHQOQOY-----OQRTLMRGGCLN-----TP 1274
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2063 KAAALAKQK-OEVRHGFQHLVSMTPARSLEFQFGLSSYINRDSIGATNGVHLHTQ 2121
Oy 1275 SMVAPSGPATMSNPRTIPOANAQGFPPRPNGISQGF-DPFGTATPQ-----SPL-MSF 1328
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2122 SIQRPSTASTISSNP--PKA-----PIQPSASVNMONTIDPAIEIRVORWFYKPLKMSA 2174
Oy 1329 RMAHT-----QSPMQOQSOANPAYQAPSDI----- 1353
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2175 EEAFTVMAVAVSSDPNPATSTVDLAAMLQLOQAQAQAQAVPVVTTASTPPLSLNET 2234
Oy 1354 ---NGMAQMG------NSMFQOQSPPHFGQOANTSMYNNMNTVSMATNTG 1399
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2235 LLASTASLANLATGALNPLSMALTTSSILNOSPYQO-----IARVLLTMNNGOMLATHOT 2290
Oy 1400 G--MSNMONTGQISMTSVTSGLSSMGPEOVNDPAL--RGGNLFPPNLPDGMXKQ 1455
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2291 SELLATYMNQOETLMALLA-----ARKGLPFAFPQONQOPOMPAQGGFALPYVLPMSLKRN 2346
Oy 1456 EGD 1458
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2347 AKD 2349
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 20
JC7771
hypoxia inducible factor-3 alpha - human
C:Species: Homo sapiens (man)
C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 01-Feb-2002
C:Accession: JC7771
R:Hara, S.; Hamada, J.; Kobayashi, C.; Kondo, Y.; Imura, N.

```

```

Biochem. Biophys. Res. Commun. 287, 808-813, 2001
A:Title: Expression and characterization of hypoxia-inducible factor (HIF)-3alpha in
A:Reference number: JC7771; PMID:11573933
A:Contents: kidney
A:Accession: JC7771
A:Molecule type: mRNA
A:Residues: 1-667 <HAR>
A:Cross-references: DBJ:AB054067
C:Comment: This protein is a heterodimeric transcription factor that belongs to the b
lved in the regulation of hypoxia-inducible gene expression in human kidney.
C:Genetics:
A:Gene: hif-3alpha
A:Map position: 19
C:Keywords: kidney

Query Match 3.5%, Score 270; DB 2; Length 667;
Best Local Similarity 20.6%, Pred. No. 5.5e-06;
Matches 172; Conservative 103; Mismatches 271; Indels 288; Gaps 37;

Oy 13 RAFTKRECEPDQGPSKRNTEKRNQENKIYELAE-LIFANFNIDINFNKRPCKA 71
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 7 KSTTELRE-----KSROAARSRQETEVYQIALHTLPFAR-----GVSAILDKAS 53
Oy 72 ILKETVQIRQIKQEKAAANIDEVQKSDVSSSTGQYIDDALGPMMLALDGGFFVYN 131
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 54 IMRLTISTLRM---HRLCAQEMNOY-----GAG---GEPLDACLKALBGFVWLT 99
Oy 132 LEGNVVSENVTOYLRYNOBELMKNSVYSILHVGDTHEFYKNLLPKSIVNGSNGSEPP 191
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 100 AEGDMAYLESNVSKHLGISOELIGHISIFDFIHPQDQELDALTPQ----- 146
Oy 192 RNSHTFCRLMYKPLPDSSEEDHNDQRAHQYEMQCFANVSQKSIKEBEDIQ----- 246
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 147 ---QTLRRRVEAP-----TERCFSLRMKSTLTSRQRTNLKNAAT 183
Oy 247 -----SCLICVARVP---MKRPVLPSESEPTT 272
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 184 WKVLNCSGMRAYKPPAOTSAGSPDSEPPLOCLVLICELAIPIHGSLEPL--GRGAFIS 241
Oy 273 RODLOGKITSLDTSTMRAMKPM--EDLYRRCIQKFAHQEGESVYAKRHHEVLRQG 330
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 242 RNSIDMKFTYCD--RIAEVAGYSPDDLIGCSAYEYTHALDSAVS---KSIHTLLSKG 295
Oy 331 LAFQIYFSLSDGTVAQAQTKSLIRQTNQOLVLSHML--HRQNVCVANPDLTG 388
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 296 QAVTGQYRELFARSGGYLMTQFQATVNSGGRGPOSEIVCFHLLSYOETGVVLSLQTE 355
Oy 389 QTMGKPLN---PISSNSPAHQALCSGNPGQDMTSSNINFPINGPKQMGPMGRFGSG 445
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 356 QHSRRPIQRGAPSQKDP-----NPGDSL----- 379
Oy 446 GMNHVSGMQATTPOGSNVALKMSNPSSQSGPMNPGOPTMLSP--RHRMSP-----GVAG 498
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 380 -----DTP-GPRLIAFLHPPSLSEAL--AADPRRFFCSPDLRLRGLIILGASVAA 427
Oy 499 SPRTPGQFPAGLSH-----PVGCSSSTGNHSHSTNS----- 532
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 428 TPSTPLATRRHPOSPLADLPDELPG---TEVNHRLFTSGKDTAEVETDIDLDAQADAL 483
Oy 533 -----SINALQALSEGHGVSLGSSLASPDLMGNLON--SPVNMNPPPLSK 576
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 484 DLEMALPYISMDDPOFLNASQQLPRAVHRLPG---AVPRPARSFHGLSPALPEPLLR 540
Oy 577 MGSIDSKDCGEGLYGPSEBGTGQAESSCHGEQKETYNDPNI--PRAVSSERADQSRILND 634
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 541 WGSDDPRLSC-----SSPSRG-----DPSASPMAGARRKTLAQSSEDE 578
Oy 635 SKGQTKLLOLITTKSDQMEPSP-----LASLSDPTNKNDSTGSLPGSSSTGTSLKEKHKI 689
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 579 DEG---VELLGVPRPPRSPSEHENTFLRPLLSIFLITGGPAGS----- 620
Oy 690 LHRLLQDSSSPVDLAKLTAEATGKDL5-----QESSSTAAGSEFTIKQEVSPK 738

```



A:Cross-references: EMBL:U22431; NID:9881345; PIDN:AAC50152.1; PID:9881346  
 A:Notes: parts of this sequence were confirmed by peptide sequencing  
 R:Henesach, J.B.; Chan, W.K.; Carver, L.A.; Bradfield, C.A.  
 submitted to the EMBL Data Library, June 1995  
 A:Reference number: H00692  
 A:Accession: G01875  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-826 <HOG>  
 A:Cross-references: EMBL:U29165; NID:91144012; PIDN:AAC51210.1; PID:91144013  
 C:Genetics:  
 A:Gene: GDB:HIF1A  
 A:Cross-references: GDB:512229  
 A:Map position: 14q21-14q24  
 C:Keywords: heterodimer

Query Match 3.5%; Score 269.5; DB 2; Length 826;

Best Local Similarity 19.3%; Pred. No. 7.9e-06;

Matches 186; Conservative 150; Mismatches 337; Indels 289; Gaps 41;

OY 1 MSGKEINTSDPSRATRRKKECPDQIGSPKNTKRNREONKYTEELAEILIFANFNDI 60  
 DB 1 MEGAG-GANDKKKISSERKE-----KSRDAARSRKSESEVEYELAHQLPRHNVS 51  
 OY 61 DNFNFKPKCALIKETVKQIRKOEKAAANIDEVOKSDVSTGGQVIDKDALGPMWL 120  
 DB 52 SHL-----DKASYMRLITSLR-VKRLDAGDUDIEDDKAKQNNC-----FYL 93  
 OY 121 EALDGFVVNLGNNVSENVTOYLRYNOBELMKNKSVYSLIHVGDHTEFYKNLLPFSI 180  
 DB 94 KALDFFVAVLFDGDMYIISDVNKNYMGLOFELGHSVFDETHCDHEMEKMLTHR- 151  
 OY 181 VNGSGWSEPPRRNRSHNTNCRLVYPLRDSSEGHNDQEAHKYETMOCFAVSQKSIKE 240  
 DB 152 -NG-----LVK-----KGEQNTQSRFRLMCTLTISRRTNMI 184  
 OY 241 EG-----EDLQSLICAVRVPKKEKRVLP-SESEFTT 272  
 DB 185 KSATKVLHCHGHINHYDTNSNQPGYKRRPMCTLVLCERIPRPSNTEILDSKTFIS 244  
 OY 273 RODLOGKITSLDTS-TMRAAMKPGMEDIVRRCIOKFAHOGESVSYAKRHHNEVLROGL 331  
 DB 245 RHSLDMKFSYCDERITELMGYEP--EELLGRSIEYEU--HALDS-DHLTKTHDMFTKGQ 299  
 OY 332 AFSQIYRRSLSDGTLYAOTKSKLRSQTNRPOLVISLH-----MLAEQNVNVC 380  
 DB 300 VTTGQYRLAKRGYVWVETQATVLYTNKNSQPOCIVCNVYVSGIIOHDLIFSLOQTEC 359  
 OY 381 VMNP-DLTGQTFMGKPLNPLISSNSPAH-----QALCSGNPGQDMTLSSINFPINGP 430  
 DB 360 VLKPESSDMKMTQLETVESEHDTSLSFDKLKKEPDALTLAPAGDTIIS-LDF----- 413  
 OY 431 KEQMKPMGRFGSGGSMNHVSGMAQTTPQGSNTYALKMNSPSSPGMNPQOPTSMLSPRH 490  
 DB 414 -----GSDNETDQQLAEVPLVN---DVMLPSNE 441  
 OY 491 RM-SPGVAGSPRIPPSQSPAGSLSPYGVCSSTGNSHTSSNALQALIS-EGHGYSL 548  
 DB 442 KLQNTNLAMSP-----LPTAETPKPL-----RSSADPALNOEVALKLENPESL 485  
 OY 549 GSSLASPLDKMGNLQNSPVNMNPPRLSKMGLSKDCFGILGPESEGTGGAESSCHGE 608  
 DB 486 ELSTFMQIDQ-----DQTP-----SPDGSSTRQ--SSPEPNS 515  
 OY 609 QKE-----TNDPNLPRAVSSERADGOSRLHDSKQTKL-IQLLT-----T 647  
 DB 516 PSEYCFYVDSDMVNEFKLELVEKFAEDTEAKNPSTQDITDLMLAPYIPMDDFOLR 575  
 OY 648 KSDQMEPSPLASSLSDTKDSTGSL-----PSSGSHGSLKEKHILH----- 691  
 DB 576 SFDOL--SPLESSSASPEASPSQSTVTVFOQTQIQEPTANATTTATYDELKTVKDRME 633

OY 692 --RLDODSSPYDLAKITPAEAT-----GKDLQSOESSSTAGS----- 726  
 DB 634 DIKILLASSPPIHINKETTSATSSPYRDTQSTASPNAGKVIQOTEKSHRPSRNVLSV 693  
 OY 727 ----EVTIKQEPVSPK--KKNALLRYLLDKDDT--KDIGLEIPPKLERLDSKTDPSAN 778  
 DB 694 ALSQRTTYDEELINPKIALQNAQRRKMEHDSJLFOAVGIOTL-----LQOPDHAAT 747  
 OY 779 TKLIMKTEKEMSFEPDQPSSELDNLEILLDL-----ONSOLPOLFP-DTRPG 828  
 DB 748 TSL----SWKRVKGCSSSEONGMEQKTIILLPSDLACRLLAGSDMESGLPOLTSYDCEVN 803  
 OY 829 AP 830  
 DB 804 AP 805

RESULT 23

JC5405

brain and muscle Ah receptor nuclear translocator-like protein, BMAL1b - human

C:Species: Homo sapiens (man)

C:Date: 10-Jun-1997 #sequence\_revision 18-Jul-1997 #text\_change 01-Dec-2000

C:Accession: JC5405; JC5404

R:Ikedu, M.; Nomura, M.

Biochem. Biophys. Res. Commun. 233, 258-264, 1997

A:Title: cDNA cloning and tissue-specific expression of a novel basic helix-loop-helix

site usage.

A:Reference number: JC5404; MUID:97289529

A:Accession: JC5405

A:Status: nucleic acid sequence not shown

A:Molecule type: mRNA

A:Residues: 1-626 <IKR>

A:Cross-references: DBJ:AB000812; NID:92094734; PIDN:BA19935.1; PID:92094735

A:Accession: JC5404

A:Status: nucleic acid sequence not shown

A:Molecule type: mRNA

A:Residues: 'MINI', 48-415, 'P', 417-626 <IKR>

A:Cross-references: DBJ:D89722

C:Comment: This protein plays a role in the control of central nervous system develop

C:Genetics:

A:introns: 47/2; 223/1; 250/2; 274/1; 288/2; 361/1; 392/1; 442/2

Query Match 3.5%; Score 269; DB 2; Length 626;

Best Local Similarity 24.3%; Pred. No. 5.6e-06;

Matches 142; Conservative 82; Mismatches 230; Indels 130; Gaps 27;

OY 35 EKRNREONKYTEELAEILIFANFNIDNFNFKPCALIKETVKQIRQIK-EQEKAAAN 93  
 DB 81 EKRNRDKMNSFLDELASLV---PTCNMMSRKLDKLVLRMAVQHMRLRGATNPTTEAN 136  
 OY 94 IDEVOKSDVSTGGVGVIDKDALGPMLEALDGEFFVNVLE-GNVVSENVTOYLRYNOE 152  
 DB 137 YKPTFLSD-----DELKHLILRAADGFLFVVGCDRGLILFVSESVFKILNYSQN 185  
 OY 153 ELMNSVYSILHVGDTHEFVKNLPLKS-----LYNGSGWSEPPR 192  
 DB 186 DLIGOSLFDYLPKDIAR-VKEQLSSDTPAPRERLIDAKTGLPVKTDITTPGSRILCSGAR 244  
 OY 193 RNSHTFCRMILV-KPLPDSSEEGHNDQEAHOKYETMOCFAVSQPSKIR----- 239  
 DB 245 R--SFCRMKCNRSVAVEKDPSTCSKKKADKRSCTJHSTGYLSKSWPTKMGLED 301  
 OY 240 ----EEGEDLOSLICVARR---VPMK-ERPVLPSSESFTTRDLOGKITSLD-TSTM 289  
 DB 302 NEPDNEGNTL-SCIVAIRGLSHVVPQVNGEIRVKSMEYVSRHAIDKCFEYVQRAFAI 360  
 OY 290 AAMKRGWEDIVRRCIOKFAHOGESVSYAKRHHNEVL--RQGLAFSQIYRFSLSDTGLY 347  
 DB 361 LAYLP-QELIGTSCYEYHQ-----DDIGHLAECRHQVLOTREKKT-TVCYAFKIDGSEFT 414  
 OY 348 AAQTSKSLIRSQTNPEPOLVISLHMLHREOVNVCVNPDLTGQTMGKPLNPLISSNSPAHQA 407



```

Db      415 TLRSRHFSPMNPWKREVEIVS-----TNTVYLVANVLEG---GDPTFPQLTASP----- 460
Oy      408 LCSGNPGDMTILSSNINFPNGPKEMGMPMGFRGSGGGMNHYSG---MQATTPOGSNTA 464
Db      461 -----HSDMSLPGSGGGKRTHTPIVPGTIRAGAGRIGRMAIEIMEIHRIRGS--- 512
Oy      465 LKMNPSQSSPCPMNPQOPFMSLSPRRHMSPG-----VACSPIRPPQSFSPAGSLHSPGV 519
Db      513 -----SPSSCGSSPLNITS-TPPPDASSPGCKILNIGTTPDIPSSGGLSGAQENP---- 562
Oy      520 CSTGNMSHTYNSLINALQALSBGCVSL-----GSSLASPD 556
Db      563 -----GYEYSDSS-----SILGENPHIGIDMIDNDGSSSPSND 596

RESULT  24
T13804
shs protein - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C:Accession: T13804
R:Treisman, J.E.; Lai, Z.C.; Rubin, G.M.
Development 121, 2835-2845, 1995
A:Title: Shortsigned acts in the decapentaplegic pathway in Drosophila eye development
A:Reference number: Z17167; MUID:96038094
A:Accession: T13804
A:Status: Preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1212 <TRB>
A:Cross-references: EMBL:L42512; NID:g833709; PID:g833710; PIDN:AAC41608.1
C:Genetics:
A:Gene: shs
A:Cross-references: FlyBase:FBgn0010460

```

[illegible][illegible]

Query Match 3.5%; Score 269; DB 2; Length 1952;

Best Local Similarity 19.0%; Pred. No. 2.8e-05;  
Matches 292; Conservative 172; Mismatches 597; Indels 478; Gaps 65;

QY 67 PDKALIKETVKQIROIKEOEKAAANIDEVOKS-----DVSSST----- 105  
Db 610 PDEDILLPADADHMEYVEKEEEDDDDEAROSQSVKPEPRPPEPSSVPLMDPPK 669  
QY 106 -GCGVIDKDALGPMMLLEADGFEFFVNVLEGNVFESENVOYLRYNOELMKNKSVYSILH 164  
Db 670 RGPGRPPKNGI-----MSKREERLKKKQAMELAK 698  
QY 165 VGDHTEPVKNLLPKSYNGSGWSEPP-----RRNSHTFNCMLVKPLPDSEEGHDNQ 218  
Db 699 -----KNQPPQP-----PGEPPVKRKYGRRKH-----PLP-----EDAPDRP 731  
QY 219 EAHOKYEM-----QCFVNSQPKSKEGEDLOSLICVARVPKKEPVLPSSSEFTTR 273  
Db 732 E-KKKYKPKRKNKEGDDSDAEKTIKEK-----RREKPKTEPLELRREDYTEE 778  
QY 274 Q-----DLOGKITSL--DTSTRAAMK-----PGMEDLVRCIOKF 307  
Db 779 QLOKPNKNYGLVIDEYLSAADGLIKOYKRIQKLPFYFENVDTKGMESSVRHNLGN 838  
QY 308 HAOHEGESVYAKRHHHEVLRQGLAFSQIYRFSLSDGLVAOGRKSKLIRSQTTNEPOLV 367  
Db 839 DAFKKN-----ETHLMSRVPIDIDAKKKR-----APSPDHA 872  
QY 368 ISLH-----MLHREQNY-CYMNPDLTQ-----TMGRP----- 394  
Db 873 SSLNFGOHVAPORPMPHPGMYHGBHGOVSQYHGTVOGROSYYTGGPQASQHPHOLO 932  
QY 395 -----LNPISNSPAHQALCGNFGQDWTLSNINFPINGKEQMGWGRFG-----GSG 445  
Db 933 TPQGVPPQOPRRARAYQAAQTSPPAQOYVGT---PPTARQMSGTPTAATYSSPYRSP 989  
QY 446 GANHVGSMQATTP---QGSNVALKMNSPQSS---PGMNGOPTSMLSPRHMSPGVAGSPR 501  
Db 990 PMPVVAOQSGATPHSMARQHSLPVGSPQANGIDPRVNPPTATA-----NTGVPAVGAR 1043  
QY 502 IPPQOFSPAGSLHSPVGY-----CSTGNSHSTNSLMAALASL-----GH 544  
Db 1044 -PAQQAATPTATTAIVRLNPVIADELISLWSEFVYVEKLEYIKTQTSKFPQILAMSVINR 1102  
QY 545 GVSGLSSIASPD---LKMGNLO-----NSPVMMNPPELSKM-----GSLD 581  
Db 1103 GKLTITSMIPDEESLELVLRVFEERQGTSHKSLDDLLQTLITFKATVSTILEALD 1162  
QY 582 SKDCEGLYGEPEEGTGOAESCHPEQKETNDPN---LPAVSSERADQOSRLHDSK 636  
Db 1163 SQKAECVLVSAIDQVLGLADKITRGTESEMEREFNNAEKVLIPAIRMKVAEQOR----- 1217  
QY 637 GQTLQDLQTLTKSDOMESPCLASSLSDTNNKSTGSLPSSGSHGSLKEKHKILHRLQD 696  
Db 1218 -----QVATTPAPVHAIR---TPGALLPANHHT----- 1242  
QY 697 SSSPVDLAKTAEATGKDLQSOESSSTAAGSEVTIKQEPVSPKKENALLRILDKDQTD 756  
Db 1243 -----MAPATAPSAQRIATPSAT-----QGRANPGCGNPSANPRKI-----TAA 1281  
QY 757 IGLPETPKLERLDSKDPASNTKLIAMKTEKEMSEFPGQPSGEIDNLEIIDLQONS 816  
Db 1282 AAPATPPALLPR---AAPISGHPRIASVTNNNTAAM---GARP----- 1318  
QY 817 QLPQLPFDTRGAPAGSV-----DKQALINDLMQLTANSNPTVPGAGAKTALRISQS 868  
Db 1319 -VPAAPGPOIGVHSAIGVATVAVSRSPVTHGVTAAPTTTPPSNLAAAVPAGSMQO 1377  
QY 869 TFNNPDRPG---QGRLLPNQNLPLDITQSPGTAGPPPIRNS--SPYSVLP--OPGM-- 919  
Db 1378 THSGAPAGSGNASISRAP-----STAPATPTVTPPVSSIVYRPASIVYPTGPPLITA 1429  
QY 920 --MGNOGMIQNGNLGNSST--GMIGNSASRPTMPGSEMAPOSSA-----VAVTCAA 967

Db 1430 PPTASSGASAGYARPANASTMPAPTSGQAAMTSAVPOQSVSPRPSSLTITGQMGVPAIAAA 1489  
QY 968 TTSMANRPVQGMIRNPAS---IPMRSSOPQOQRTTQSOVMNIGPSSL---EMMGGPO 1022  
Db 1490 STS---RPASG--VYNPASSLAPSTHKSMPSAVPTTASGAVSSVSLAATPLPSPAPR 1544  
QY 1023 Y--SQOQAPNPQATAPWPSIILPIDQASFASONROQFSSPDDLCPHPAASPSDEGALL 1080  
Db 1545 YGPSNNATPTPTATAIPIRAPLIPAASSVSAPVTGQSFSPAPASVPPPTSAAS----- 1596  
QY 1081 DQYLALRNPDLGEEIDBALGIPELVSOQAVDPEQFSSQDSNIMLEQAPV--FPQOYAS 1139  
Db 1597 -----QGAQPLSQPVGPAPAPAISSIGATPAASIPSSAPAL---APVTYVPQQA 1643  
QY 1140 QAQMAQSVSPMDQNFHTMQRPSYALIRMQPRRGLGPRGTGYQVQNPQOLQOLHRLQA 1199  
Db 1644 SAAARLPVTPA--PAATHTAOSVA-----QVAPR---VYQSP--VQSVAAQHTSS 1688  
QY 1200 QON---RQPLMNQISNVSNVNLTLRPGVP-----TQAPINAOMLAQOREILNQHLRQ 1249  
Db 1689 QASTTAHPVAAQSVPRPVNSMPTSAAPVAAVCTAQVAAAPVTYQAPRHALSSVSQSLPQ 1748  
QY 1250 ROMHQOQOQVQORTLIMRQOGLNMTPSVAPSGMAPATMSNPRIPOANNAQOFPPPVYGISQ 1309  
Db 1749 SVPHAAQQAQAHQT-----PHSASRPVQSVPOS-----VQKRTQAVVPRPSTSLTP 1795  
QY 1310 QPDGFGTATTPQSPILMSPRAHOTOSPMAOQSOANPAVOA----- 1349  
Db 1796 TAPQGPVSPAVSGSVPPAPSAQSVAPAPVASTPVPAATVAPASTVAAAPPTRYTAAPA 1855  
QY 1350 -----PSDINGMAQNGMNGNSMFQSPPHFGQOANT 1381  
Db 1856 ALSATINPAPVPSQPHQITQOAPRQOQRRPAPQAQAPAT 1894

RESULT 26  
T30556  
aryl hydrocarbon receptor alpha chain - rainbow trout  
C:Species: Oncorhynchus mykiss (rainbow trout)  
C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 11-May-2000  
C:Accession: T30556  
R:Abnet, C.C.; Tanguey, R.L.; Hahn, M.E.; Heideman, W.; Peterson, R.E.  
J. Biol. Chem. 274, 15159-15166, 1999  
A:Title: Two forms of aryl hydrocarbon receptor type 2 in rainbow trout (Oncorhynchus  
A:Reference number: 220858; MUID:99262676  
A:Accession: T30556  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-1058 <ABN>  
A:Cross-references: EMBL:AF065137; NID:g4007963; PID:g4007964; PIDN:AC95335.1

Query Match 3.5%; Score 268.5; DB 2; Length 1058;  
Best Local Similarity 20.0%; Pred. No. 1.2e-05;  
Matches 251; Conservative 170; Mismatches 449; Indels 387; Gaps 63;

QY 17 KKRKECPQOLGSPR-----KRNTERKRNQENKYEIELAELI--FANFNDINDNFNPKPKC 70  
Db 11 KKRKKPVQKTKKSPAPDVYKSNPSKRHRDRLNGELDRTLGLLPPR-----EDVSRIDLK 65  
QY 71 AILKETVQKQIROIKEOEKAAANIDEVOKSDVSTGQGVINDKAL-----GPMMLLEAL 123  
Db 66 SVLRISVGYLK-VKSFEKTT-----MKSSVLPPGGGGLMMNGMNAATPFSEGLLLQAL 118  
QY 124 DGFEFVNLEGNVVFVSENVOYLRYNOELMKNKSVYSILHVGDHTEPVKNLLPKSYNG 183  
Db 119 NGFVLVYTAEGHVFYSSPTIDYLGAFHQSDVHVQSVFELIHTDRAPE----- 166  
QY 184 GSWGSEPPRRRSHTFNCMLVKPLPDSEEGHDQEAHQKETYMQCAVSPQSKIKEGE 243  
Db 167 -----RQJLH-----FALNPKRFDEQGDGMA--RSSDITRNIVTYNPEDLPPENS 211  
QY 244 D-IQSLCLICVARVPKKEPVLPSESEFTTRQDLOGKITSIDTSM----- 288

```

Db      212 SFLEKNVC-----RFRCILDNNSGFLA-LNFQGLKFLHAQSMIGDGTGSRQPNLGL 263
Oy      289 -----RAAMKQGMEDLVRCI-----OK 306
Db      264 FTIATPVQNPSTLEIRKTIFFQTKHKLDFTPMGVADKRCVVLYGYISM-ELCHRGSSYQF 322
Oy      307 FNAHQEGESVYAKRHHHEVLROGLAFSQIYRPSLSDGLVAAQTKSLIRSQITNEPOL 366
Db      323 IHA-----ADMMFCADNHNVRMIKTGESGLTFRLQTKGCWVWVQANARLV--YKGGPRDF 376
Oy      367 VISL-----HMLHR-----EONVCVMNDLIG-----QTMGRKLNPTSSN 401
Db      377 IARQBALLNSEEHEHLROKRMELPESFTTGLEALYETGPTLDATPEFOTSPKIRKRESL 436
Oy      402 SPAHQALCSGNPGQDMTLLSNINFPINGPKQGMPMGR-FGGSGGNHY-----SGMQATP 457
Db      437 DP--QSLTGLSMQDESITYQ-----FOERO-LPDLQAFMDSKRLTLYNACHSMQSSME 466
Oy      458 PGGSNTALKMNSPSSQSPGNPGQPTSMLSPRHRMSPGVASPRIPPSQSPAGSLHSPV 517
Db      487 PGPRPDGDDDDGDPSEVN--QKGAIVYAMIDALEKMARL-----GDL----- 524
Oy      518 GVCSSITGNSHTYNSLSNLQAL--SEGHCYSLGSSLSASDLKMGNLQNSPVMMNPPLS 575
Db      525 -----CEALQGLDVEAEELMEWESAL--LRLSQESNQTGGGDTSP-- 562
Oy      576 KMGSLSDKDF-----GLYGEPEGTTGQAESCHPEQKETNDPNLPRAVSERADQGS 630
Db      563 ELDDITNDNDFSYVEELHFNESSEGSQNPNCST--MYNNPN----- 603
Oy      631 RLHDSKGQTKLQLLTTKSDOMEPPSLASSLSDTNKSDTSLPGSGSTHGT-----SLKE 665
Db      604 -----LNLFT-----EVFNNNNOD--GPPGMVSPGTGVCCKPGLLD 638
Oy      686 KHKILRLIADSSSPVDLAKTLAATGKDLDSQSSSTAPSEVYTIQEPSPKKKEAL- 744
Db      639 SRSFTI-----NSPVN--SLNGOVTGN-----GPDLAGCQNAQPHQVFESTQ 660
Oy      745 -LRYLLDKDQTKDIGLEI-----PPLERLDSKTPDPAATKLIAMKTEKEMSFE 794
Db      681 RLSHFGPQIPQMDLNTPTLQQLNDLIFTPSL-LPELSLPHSSGNGAV-TCTNNAGS 738
Oy      795 PGDQPSSELDNLEILLDLQNSQLPOLFPDTPRGAP--GSVDKQAI-----I 840
Db      739 CAQAPRNHMSPOGITGRVHNSQPPQF--FTNHGLPATMASNGPQISVQSNHVAPNLV 797
Oy      841 NDMLQTLAENSPTPYGAQKTALRISQSTFN--NPRPGQL--GRLLPNQMLPDLITLQSP 896
Db      798 DGNASMLPSPNAFVSP-----QIESSNINLSNPLPTACLOGNSAFQSLKIQRYLQWP 849
Oy      897 TGAGPPPIINSSPYVYIPQGMNGMGNOGMLGNSSTGMLGNSASRPTMP----- 950
Db      850 QNQOQLRP-----PASTI-QNGIMANGHTIPIRCHSDSTQRYVPLGLIMPQPNRLXHQ 903
Oy      951 -----SGEMAQSSAARVYTCATTSAMNRPVQGMIRNPAASIPMRPSSQPGQROTLQ 1003
Db      904 TQHGGLANCGPAPSSSC-----MFENISPHLPLNGNSHVDGTR--LA 942
Oy      1004 SQVMNITGPSPLEMMNGGPOXSQQAAPPNQA--PWPESTLPIQASFAASNNRQPFSSPD 1061
Db      943 STL-----SYCQSRMWDP--QDOSPPKSGCYQWQGPS-EPVVGTSVAVIDST--STSP 991
Oy      1062 DLICPHRP-AAESPDEGALLDQLYLALRNFDGLEI-----DRAIGPELVQS 1109
Db      992 S-----RPVANTITTEGILLAMQOYLACGSCVGTQIPLSPVIDSNGILSPPLVNGS 1044

```

C:Accession: T00022  
R:Takeuchi, T.; Chen, B.K.; Qiu, Y.; Sonobe, H.; Ohtsuki, Y.  
Gene 204, 71-77, 1997  
A:Title: Molecular cloning and expression of a novel human cDNA containing CAG repeat  
A:Reference number: 214059; MID:98094256  
A:Accession: T00022  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-1142 <TAR>  
A:Cross-references: EMBL:AB001895; NID:g2588990; PIDN:BA423269.1; PID:g2588991  
C:Genetics:  
A:Map position: 1p35-36.1  
A>Note: B120

Query Match 3.5%; Score 268; DB 2; Length 1142;  
Best local similarity 19.6%; Pred. No. 1.5e-05;  
Matches 262; Conservative 116; Mismatches 424; Indels 536; Gaps 61;

```

Oy      382 MNPDLTGQTKGRKPLNPSSNS--PAHQALCSGNPGQ-----DMTLLSNINFPING 429
Db      7 MRQPYRGCT-----NPSQQQGGPSDPDQGHGVRGQPYGSGTRQRYRMVYQGRAGSAMG 61
Oy      430 -PREQGMPEGRFGSGGGMNHVSGMQATTP--QGSNTALKMNSP-SQSSPGMNP-GQPT 483
Db      62 LSYTQOIPRYGQGGQPSG--YGOOGQTPRYNQOSPHPOQOQPPYSQOPPSQTPHAPGS 116
Oy      484 SMLSPRHR-----MSPGVAGSPRIPPSQSPAGSLHSPGVCSITGNSHTYNSLSNL 537
Db      117 YQOQPOSQPPQLOLSSQPPYSQOPSPRHQSPA-----PPSQOSTTQOHP----- 162
Oy      538 QALSECHGVSLGSSLSASDLKMGNLQNSPVMMNPPLSKMGLSDKDFGLYGEPESEGT 597
Db      163 -----QSQPP-----YSQP----- 171
Oy      598 GQAESCHPEQKETNDPNLPRAVSERADGGRINDSKQGTLLDQLTKSQMPEPSL 657
Db      172 -QAQOS--PYQOQOPOQP-APSTLSQQAAYPQOSQSQ-----QTAYSQOQREPP- 217
Oy      658 ASSLSDTNKSDTSLPGSGSTHGTSLKEKHKILRLDQ-----SSSPVDLAKTLAE 709
Db      218 -----QELSDQSPGS-QASSAPMTSSKQGEQEDMNLISQSRPSSLPLDLSGIDDLPGCTEG 272
Oy      710 ATGKDLSQESSSTAPSEVYTIQEPVSPKKKEALLRYLLDKDQTKDIGLEITPXLERL 769
Db      273 ALSPGVSTSGISSQGEQSNPAQSPSPHNSPH-----LPGI-----RG 311
Oy      770 DSKTDPASNTKLIAMKTEKEMSFERGDQ-----PSELD-----NLEIILD-----L 813
Db      312 PSPSPVGSPPASVQASRGLSPPAVPPGNOMPPRPSSGSSDSIMHPSMNOSSIAQDRGYMO 371
Oy      814 QNSQLPOLFPDTPRGAPAGSVDKQAIIINDMLQTLAENSPTPYGAQKTALRISQSTFN-- 871
Db      372 RNSQMFQ-YSSPPGSGAL-----SPQLSGGQ---IHTMGSTYQON 408
Oy      872 -----NPRPGQLRLPNQMLPDLITLQSPGAPPPPIINSSPYVYIPQGMNGMGMI 926
Db      409 SMGSGPGQGOYG-----PQGYPRQPNYNALPNANYPSAGAGGINP 452
Oy      927 GNGGNL-----GNSGTGIGASARPTMP--SGEVA----- 955
Db      453 GAGGQNHGQGPPIPYGTLPBGRSHASMGN--RPGYPNNGQYATSGWWDVSPPGGMNR 509
Oy      956 -POSSAVRYTCATTSAMNRPV-----QCGM-----TRNPAASI-PMRPSSQPG 997
Db      510 KTQETAVAMHVA--NSTQNRPPGYPMNMGCGMAGTGPPRYGQGINSAAGMNPQGPYYSMG 568
Oy      998 QROTLQS-----QVMNITGPSEL--EKNM--GGPOXSQQAAPPNQAAPPWPESTLPI 1043
Db      569 GTMANNAGMAASPENMGLDVKLTPTATKMNKKAADGTPKTESKSSKSSSTTNEKITKL 628
Oy      1044 -----DQASFASQNRQPPRGSSPDLLCPHPAAESPDEGALLDQL 1083

```

RESULT 27  
T00022  
B120 protein - human  
C:Species: Homo sapiens (man)  
C:Date: 22-Jan-1999 #sequence\_revision 22-Jan-1999 #text\_change 21-Jul-2000

Db 629 YELGGPERKMWDRYLAFTEEKAMGNTNLPVGRKKPLDLY-----RL 671  
Qy 1084 YLALNPFGLSEIDRALGPEL-----VSSQAVD 1113  
Db 672 YVSAVEIGLTVQVKNKRWRELATNLNGTSSAASLSLKQYIOCLYAECKIERGEDPP 731  
Qy 1114 PEQFSODSNIML-----EOKAPVEPQOYASQAQMAQ----- 1146  
Db 732 PDIFAAASKKKSQPKIOPSPAGSGSMGQPPQ-----STSSMAEGEDGLKPPRPASTPHSQ 789  
Qy 1147 -----SYSPMDP 1154  
Db 790 IPELPGMSRSNSVGIODAFNDGSDSTFQKRNMTPEPXPMSMTSDMMGRMSEYENKDP 849  
Qy 1155 -----NEHTMGORPS-----YATLRMQPR-----P 1174  
Db 850 YGSMRKAPGSDPPMSSGGGPPMGMDPYSRAGPELGNVAMGRPHYRGCPHYDRVTEP 909  
Qy 1175 GLRPTG-----LVONQPNQ-----LRLQLOHRLQAQONRQPLNMQISNVGNVLT 1219  
Db 910 GIGPEGNSTGAPQSNLMPNSPDGMSYPSRKYPPQOQOQOQOHRDSTGNQEST----- 962  
Qy 1220 LRPGVPTQAPLINAQMLAQRREILNQHLLRQROMHQOQOYQORTLM----- 1264  
Db 963 --QGTPSGSPSPSQ-----QTTMYQOQOQOQNTKRPMDGTGYPAPKRRHEGM 1005  
Qy 1265 -----MRQGLMTPTSMVAPSGMPATMSNPRTIPQANAOQFPFRPYGIGSQDPDPEGTAT 1319  
Db 1006 YSVPTSTGQLPQOQOL--PPAQPPASQPPQAQPSPOODVY--NOYGNAY--PATATPA 1059  
Qy 1320 TPQSPFLMSPRAHQSPMAQOQSQANPAYQAPSDINGMAQNGNMGNSMFSQ--OSPHEFGQ 1378  
Db 1060 TERRAGAPQ--NQFPF--QFGDRYSAPPGCTN--AQQNPMPQMGPIQASAEVVAQ 1111  
Qy 1379 ANTSMTYSNNMNINVSMTATNT 1398  
Db 1112 GTMNGRNDMTYNYANROST 1131

RESULT 28  
S26058  
Probable transforming protein (can) - human  
C:Species: Homo sapiens (man)  
C:Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 08-Oct-1999  
C:Accession: S26058  
R: von Lindern, M.; Fornerod, M.; van Baal, S.; Jaegle, M.; de Wit, T.; Buijs, A.; Grosve  
Mol. Cell. Biol. 12, 1687-1697, 1992  
A: Title: The translocation (6;9), associated with a specific subtype of acute myeloid le  
k-can mRNA.  
A: Reference number: S26058; MUID: 92195315  
A: Accession: S26058  
A: Status: preliminary  
A: Molecule type: mRNA  
A: Residues: 1-2090 <YON>  
A: Cross-references: EMBL:X64228; NID: q29652; PIDN: CAA4553.1; PID: q29653

Query Match 3.5%; Score 266.5; DB 2; Length 2090;  
Best Local Similarity 20.7%; Pred. No. 4.1e-05;  
Matches 345; Conservative 181; Mismatches 600; Indels 541; Gaps 84;

Qy 1 MSGGENTS-----DPSRAETRRKRCPCDQGLSPKRNTEKRRREQENKYEIELALIFAN 56  
Db 703 MAGIGEELAHQKLELEKARTSKAC-FQVGTSEMKMLRTESDLHFLLEIKETTESL 761  
Qy 57 FNDIDNFKPKDCAILKETVKQIQIKOEKAAANIDEVOKSDVSTGQ--VIDKD 113  
Db 762 HGDISS-----LKTITLE-----GFAVEEARQENRRNDGSLHLTKR 801  
Qy 114 ALGPMLALDGGFFVVLVEGNVYVSENYOYL-----RY----- 149  
Db 802 PLDPKSEAIOE--IRRLHQYVKFAVDVNDVLDLEWDLQKOKKORHLLVPERETLF 858

Qy 150 ----NOEELMKNYSVILHVGDTPEFVKNLLEPKSIYNGSW----SGEPPRRNSHTFN--- 199  
Db 859 NTLANNRELIINOQRRLNLHVDLSLOQLRYKOTSL-----WLSAAVPSQSIHSFSDOL 913  
Qy 200 ---CRMVLPPLPDEEGHNOEA----- 220  
Db 914 ESLCNALKTETTESHTKSLPKVPAKLSPMKQAOLNRLAKRKTPPVRSAPASLSRAFL 973  
Qy 221 -HQKETMQCFVAVSQPKSIKE--EGEDLQ-SC-----LICVARRPKMERP-----VLPS 266  
Db 974 SQRYEEDL--EVSSTSSVSQSLSEEDARTSCDEADVAQARPAVYVATPPIOSPILPH 1031  
Qy 267 SESFTTODLOQKITS-LDTSTMRAMK--PCMEDLVYRCIOFPAOHEGEVSYAKRRH 323  
Db 1032 AAPFAKSHLVHSGSSPCWCTSVATSAKIIPOGADSTMLATYT--VKHGAAPS-----H 1084  
Qy 324 HEVLROGLAFSQIYRPSLSDGTLVAQTKSKLIRSQTNEPOLVLSLHMLHREQNYCV- 382  
Db 1085 PISAPQQLAALALRRQMASQAPAVNTLTESTL-----KNVPQV-----NVOELK 1129  
Qy 383 -NPDLTGOTMGRPLMPISNSPAHQAL--CSGNPQODMTLSSNINPPIGPKQMGKPMG 439  
Db 1130 NNPAATPSTAMGSSV--PYSTAKTPHPVLPVPAANOAKQGLSLNSLK--PSGTPPASQ-LS 1185  
Qy 440 RFGSGGMNHVGMQATTPQGSNTYALKNMSPSSPGMNP-----OPTSMLSPRHRMSPG 495  
Db 1186 SCDKASGTAKITTAVTSTPSASGQFSKPPSTSPSGTGFNGITLTPPSNFTAAAGATPS 1245  
Qy 496 VAGSPRIPPSOPSPAGSLHSPVGVCSSTGNSHSYTNSLSNALQALSEGHVSLGS----- 550  
Db 1246 TKESSQ--PDARSSG-----GSKRSY-----EALPSSPSGITSASNTTPG 1286  
Qy 551 ----SLASPDLMGN-LQNSPVNMPPLPSKKSLSDKCFCGLYEPSEGT----- 597  
Db 1287 EPMASSSRVAPSGTALSTTSKLETPP-SKIGEL-----LPSSLAGETLGSFSGLR 1338  
Qy 598 -GQAESCHPGEOKE-----TNDPNL-----PPAVSEERADQS 630  
Db 1339 VQADSDTKPTNKKASTSLSTQPTKTSVSGFNFTAPVYLGKHTPPVYSS----- 1391  
Qy 631 RLHDSKQTKLQLLTKRSDQMEPSPLASLSLDTNKDSTGSLP-----GSGSTHGTSL 683  
Db 1392 -----ATTSS--VAP--PAATSTST--AVFGSLPYVASGSGVTSFGTSL 1432  
Qy 684 KEKHILRLLODSSPVDLAKLTAEATGKDISQSSSTAPGSEVITIKOEPVSPKKENA 743  
Db 1433 S-----AKTISFSG--SQQTNSTVP-----PSAP----- 1455  
Qy 744 LLRYLLDKDDTKDIGLPEITPRL--ERLDSKTDPAASNTKLAMTEKEKMSFERDQPG 800  
Db 1456 -----PPTTAATPLPTSPFLSGLSLSAATTP--SLPMAGSTEBATSSALPERAG 1506  
Qy 801 -SELD-NLEELIDLQNSQLDLPDTRPGAPAGSVDRKQAIT-----NDLM 844  
Db 1507 DSEVASAASLIEEQQSQALQPPAPQT-----SDSVKEPEVLAQPAVNSGTAASSTSLV 1561  
Qy 845 QLTAEKSPVTP-VCAQKTAALRISQSTFNNPBPQGLGRLLPNONLPLDITLOSPTGAGCP- 902  
Db 1562 ALSAETPATGCVPARTEAVPPASSFS--VPGQTA-----YTAALISAGCYVA 1608  
Qy 903 -----PIRNSSPYVITOP-----GMGQMGIGNNGNLGNSSTG--MIGSASRP 947  
Db 1609 VETSSPTLASSSTTSIYVAGPSPAAAFGTVYSGSVFQPPAASSSSAFNOLTNNTAT-- 1666  
Qy 948 TMPGEMAPSSAVRVTCAATTSAMNRPVGGMIRNPASLIPRPS-----OPG 997  
Db 1667 -----APSATPVPGQYAASTAPSLFGQQTGSTASTAAAPQVSSGFSAPFTAPG 1719  
Qy 998 Q-RQTLQSVANIGPSELE---NMAGPQYSQQA--PPNQTAPWPSILPIDOASTAS 1050  
Db 1720 VFGQTTFGQASVFGQSASSASVFSFSQPFSSVPAFQOPASSSTPTSTSGSVFQAASST 1779  
Qy 1051 QNRQ-PFGSSPDLLCPHPAASPSDEGALLDQLYL-ALRNFDGLEIDRALGIPELVSO 1108





```

Db 106 LLEGSTQOLVDPMRPGSGSFFVAGSHR--OSQLPSSHASILLPEBELGIPKVEVPRPSQV 164
QY 212 -----EGHNDQNAHQKLETMOCFANVSOPKSIIEEGEDLOSCILICVARRPVK----- 259
Db 165 SLKPAEAAHKKKPKPKGKYICQYCSR-----CAKPSVLQKHINSHTG 209
QY 260 ERPLV--PSSESEFTTRDRL-----OGKITSLDTSTMRAMKPGWEDLVRRCIQKHAQ 310
Db 210 ERPYPCGCGCFKTKNLKHKRSHAHRIKAGLASGSSEMYPRGLEMERIGEEPEEP 269
QY 311 HEGES-----VSYAKRHHEVL---RQGLAFSQIYRFSLSDGLVAAQTCKSLINSQT- 360
Db 270 TEGESTSEETGAASPSVDLPRKHPRLSSSLY-----SSGSHGSGSQERCSLSQSTG 325
QY 361 -----TNEPOLVISLHMLHREONV---CVNMPDLTGOTMG- 392
Db 326 PSLIEDPAFPAEASSENPLSHKPEDTHTIKOKLALRLSERKKLIEEQTLSPGSKGTESG 385
QY 393 -----KPLNPISNSPAHQALCSGNPGQ-----DMTLSSNIN-----F 425
Db 386 YFSRSESAEQGVSPPTNNAASYAEIIFGKGRIGQRTSMLASSTQPLPLLSSEDRSLV 445
QY 426 PINGPKQMGMPMGREGSGSGMNHV-----SGMQATTPOGNYALK--MNSP 470
Db 446 PLVSPRTQV-----IEHTIKLITINEAVVDTEISIDSVKPRRSSILTRRSVESP 493
QY 471 -----SQSSPGMNPQPTSMLSPRHRMSPGVAAGSPRIPQSPAGSLHS---PYGVC 520
Db 494 KSLIYRDSLSHGEKTKOEOBSLJSLQH-----PPSSHVPPLRLSHSMPSAAC 541
QY 521 SGTGNSHSTYNTSSLNALQALSEHGVSIGSSLASPDLMKGNLQNSPVNMMNPPRLSKMGL 580
Db 542 TISTHHTTFRGS-----YSFDDHADVPS---RMTPVFTSHPRMLKRNAA 585
QY 581 DSKDCFLYGPBGTGQAESSCHPEQKETNDPNLPRAVSSRADQGRSLDSKG-QT 639
Db 586 IELPLGGEY-----SSEEPG--PSSKDPSTKPSDEPPKESDILTKTKKPKFT 631
QY 640 KILQILTLT-----KSDQMEPSP--LASSLDTNKDSTGSLPESGTHGSLKEKA--- 687
Db 632 KGANYECTTIGARKKKKNDNTBAHKKYICSELOITKAHSVGA-----HEVEKTAQAPPP 684
QY 688 --KILHRLLODSSSPVDLAKLTAEATGKDLISOESSSTA--PGSEVITIKOEPVSPKKR-- 740
Db 685 WSGQMNHKL--GATLTLTPLRKRRKKESLGDDEEPAPACPGSETAHNRPIGSTSPA 741
QY 741 ENALLRLLDKDITKIGLP-----EITPK---LEPLDS--KTDPASNTKLI 782
Db 742 EASKASAPSLIEDPRASSPGLSQELQGNQGRGECQPKFTVIOHTSSFEKSDPEQPSGL 801
QY 783 AMKTEKREMSPEPDQGSSELNLEETLIDLQNSQOLPFL-----PDRRPGAPASVD 835
Db 802 EEDKPPAOFSSPPAPRGRSAHSIQPRLVROPNTQVPELILTEEPDRDPEPEPEPE 861
QY 836 KOALIT-----NDLMQLTAEKSPVTPVGAOKTALRIS-----OSTFNNRPGQLRL 882
Db 862 KTEEFQMPQRTQTLQALPAELLP-----PKKRLRLAEMAQSSSESSFESEVP--LBR-S 913
QY 883 PNQMLPLDITLQSPGTA-----GPPPIRN---SSPYSLVP----- 915
Db 914 PSQSSSISLSGSSRSASAFREDHGAKAEPGFSDTRSKTLGSHMLTVPSSHPARREMRS 973
QY 916 ----QPMMGQMGIGINOG--NLGNSSTGMIGNSASRPTMSGEMAPQSSAVRYCATPTS 970
Db 974 ASQSPVAVPHSHMTETRSKSFYDGLSPGSLAVPAAPPPAPAP--RRKCLVRO 1030
QY 971 A-MNRPVQGGM-----IRNPASISPMRPSQP-----GQRQTOQOVYMNIGPELE 1015
Db 1031 ASLNRPEAELEAVRKNRAVRMPASKSPSTKSSVPOISVGTGQ-----GPS--- 1079
QY 1016 MNMGPOYSQQAAPPNOTAPWESILPIDQASFASONKROPFGSSPDLLCPHPAESPSD 1075

```

---

```

Db 1080 ---GKSQMODRPPGLSSPPYTEAL-----OVFQPLGTQ-----LPPAS----- 1116
QY 1076 EGALLDQVLYALRNFDGLEIDRALGIPELVSOQAVDEQFSSQSNIMLEQAKAPFPQ 1135
Db 1117 -----LFSIQQLLPQEQ-----QSSSEFFPT 1137
QY 1136 QVASQAMAGSYS--PMQDNEFTMGQRPSTATLRMQPRPGLRPTGLVONQPNQRLQ-- 1192
Db 1138 Q--AMAGLSSPYSMPLEPP--SLFQAP-----PLP-LQPIVL---HNSQLHPQL 1180
QY 1193 LQHRLAQONRQPLMQI-----SNVSNVNLTLRPVPPQAPINAGMLAQOREILNQHL 1247
Db 1181 LPHADIPRQPPSFLPMPCPAPSTLSGYFL-----PLQSFALQPEET-ESH 1229
QY 1248 RQRQMHQOOQOVQORTLMMRQGLNMTPSWAPSPGPMATNSN-----PRIPQANAQOF 1299
Db 1230 PPVK-----TSLPPLATGPRG--PSSSEYSSDQLQPLPPVTPQATSPAP 1270
QY 1300 PPPNTGISQOPD-----PQFTGA--TTQSPMSFRMAHTQSPMMQSQ 1342
Db 1271 TSAPLPLAPACPDPAMVSLVVPVRIQTHMPSYGSAMVTTLSQILV-----TQSPGSPAST 1324
QY 1343 ANPAYQAPS-----DINGMAQGNMGNSMFQSOOS---PPHF--GQQAATSMYSNN 1387
Db 1325 ALRKYEPSSKSMTVCEADY---YEAEPGSSISKQONRGYOTPYLVRPERKGTSLSSBG 1381
QY 1388 -NNINVSAMATNTG-----GMSNMQMTGQ-----ISMTSVTSVS 1420
Db 1382 ILLEGGSSYASGSKRYLSPAGSLLEJMETQOQKRYKEEASAKDEKLELVYSCVVLRS 1441
QY 1421 TSG-----LSSMCPROYNDP---ALRGMLFPNQLPGMMIQA 1455
Db 1442 TEDRKTEKPHVGQGRREARETSLSSD--VSDPEKLSPLSHSTLSHGTAGSBAKHE 1500
QY 1456 EGDPTTK 1462
Db 1501 YAOPTSK 1507

```

---

```

RESULT 32
T34513
hypothetical protein ZK783.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T34513
R:Favell, A.; Vaudin, M.
submitted to the EMBL Data Library, August 1994
A:Description: The sequence of C. elegans cosmid ZK783.
A:Reference number: Z21536
A:Accession: T34513
A:Status: preliminary; translated from GB/EMBL/DBJ
A:molecule type: DNA
A:Residues: 1-3507 <FAV>
A:Cross-references: EMBL:U13646; PIDN:AAC24418.1; GSPDB:GN00021; CESP:ZK783.1
A:Experimental source: strain Bristol N2; clone ZK783
C:Genetics:
A:Gene: CESP:ZK783.1
A:Map position: 3
A:Intons: 14/1; 48/2; 84/1; 196/3; 303/1; 381/1; 586/1; 605/1; 1175/3; 1207/1; 1409/
3504/1

```

---

```

Query Match 3.4%; Score 258.5; DB 2: Length 3507;
Best Local Similarity 20.5%; Pred. No. 0.00021;
Matches 232; Conservative 151; Mismatches 451; Indels 295; Gaps 55;

```

---

```

QY 360 TNEPOLVISLHMLHREONVCMNPDITGQTMCKPLNPISNSPAHQALCSGNP----- 413
Db 1984 TTAEPHTVTSI-----SSTSTKDWMT-----SSKSPENVMTSESSEVSTSS 2025
QY 414 ----GQDMLSSNINPPIGPKQMGMPMGREGSGSGMNHV---SGMQATTPOQ---GSN 462
Db 2026 SKSTTASSETTVSST-----PSSSSSEAPLTSSPATTTTEVTESSVKSTTKRESSSE 2078

```





Db 1089 LSDPRLSTTPQVAQRPV-----QTITMSNVITTPSSETPRPQSAANYQVYPMNP 1137

Qy 655 SPLASLSDTMKDSGTGLSPGSGSTHGTSLKEKKKILHRLDSSSPVDLAKTLAETGKD 714

Db 1138 SLPPSSHSESOVRSRTKLP-----SNKKED-----IRHSTCDNCKKCSSEKSNAD 1184

Qy 715 LSOESSVAPEGEVITKQEPVAPKKKENALLRYLKDQDKTIDGLPEITPPKLERLSKTD 774

Db 1185 ISMOP-----PASTIV-----TYPQSRMDQHLE 1209

Qy 775 PASMTKLIAKTEKEEMSE-----EPGDQGSSELDNLEETLDDLONSOLPQ 820

Db 1210 QQRGEQLA-ATDDQICCVKYLPEIHSHTNIGRPQSKSPSGVNGTSHGLTNGPPP 1268

Qy 821 LEPRTRPAPAGSVKQALINDLMQLA-----EASPTPVGAQKLTALRISSTFNPR 874

Db 1269 LPPPR--SRAPKPKKNAFEDDMNASSGHAAYOMQSPSEVSIK-----LLKQHLN--R 1319

Qy 875 PGQLGRLPNNPLRPDITLQSPGAPRPPRINSSPYSLP-----QP 917

Db 1320 PNOFNINISQSIDL-ATLQDSVMSLP-KLQRTSEQTQMKFSHNQKQAVHEKKNNMQ 1377

Qy 918 GMMGNQGITGNOGILGNSSTGMITGNSASRPTMPSGEMARQSSAVRYTCAATTSAMNRPVQ 977

Db 1378 GSAKQEPFHQNVNQVPTQSSGPAHQSTI--HYAEMKQIKLT-----1421

Qy 978 GGMIRNPAASIPMRSSPQCGROTLQSQVMNIGTSPSELENNMGSPQYSSQQAQRPNOTAPMP 1037

Db 1422-----SLQSNKTSQCHOSTTLQAPV-----EAPSDQ--FP 1450

Qy 1038 ESILPIDAASFASQNRPPG-----SSPDDLCPHPAAEPPSDGALD-----1081

Db 1451 ENFL--EQRPSPDKVTEVIYNNVASVLSPPDE--RGQKQFTKTEKHLIDGSKTPRK 1505

Qy 1082 QLYLALRPFDELLEIDRALGIRPELYSQSAVDRPEQSSQDSNIMLEQKAPVFPPQYASQA 1141

Db 1506 QSHVDRKRSRGSF--VHYVVGKQOMISQDQNRNAYUYOQOO-----QPNASSL 1550

Qy 1142 QMAQGSYRPMQDPNHTMGQAP-SYATLRPAPRGL-RTGLVONOPNOLRLQLD-----1194

Db 1551 STMNDQYEVLD---NTINQNPMSDEQYQVMNYQTSIVQTSVAGQRLQDQNDIQVYINQ 1607

Qy 1195 -HRLAQGNRPRLMNOISVNSVNTLTLPGRVPQADINQMLAQORE-----1241

Db 1608 THRYVQYQHN--PVQGNQNRN-GRTKAARPKRTKPRPAVQSSRSVALNHERAQIVDFAKTOPA 1664

Qy 1242-----ILNQLRQRMHQOQOYVORTLMMRGQGLNMPRSVAPSGMPATMSNPRLPOA 1294

Db 1665 DQEDQTVLQDHNENYNSQQRPEYQHRLVQME-----QS 1700

Qy 1295 NAOQ---FPPRPYGISQO--PDPEFTG-ATTPQSPILSPRMANHTGSPRMQOSQANPAYQ 1348

Db 1701 NBOQSQVFOHONHOAOQELANQENQSMGMAHQOQQRQOSQLOQLOLRPQOQMOIQOYQ 1760

Qy 1349 --APSDINGMAQGNNGNSMFSQSPRHFQQAQNTSM-----YSNNMNTNVSM 1394

Db 1761 NVGSHIQ---KQNAAPQAV--QOQPMQYAAQOQOQPMQAMQADQSRQOEQVYQGOVQORTG 1815

Qy 1395 ATNTGKMS--NMQMTGQISMTSVTSVST 1421

Db 1816 ATHTVLTSPRMNHHSQANVLRLVAPKOT 1844

RESULT 34

T13998

gene mastermind protein - fruit fly (Drosophila virilis)

C:Species: Drosophila virilis

C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 17-Nov-2000

C:Accession: T13998

R:Newfield, S.J.; Tachida, H.; Yedvodnick, B.

J. Mol. Evol. 38, 637-641, 1994

A:Title: Drive-selection equilibrium: homopolymer evolution in the Drosophila gene master

A:Reference number: Z17850; MUID:94365848

A:Accession: J13998  
A:Status: preliminary; translated from GB/EMBL/DBD  
A:Molecule type: DNA  
A:Residues: 1-1655 <NEW>  
A:Cross-references: EMBL:M92914; MID:9157833; PID:9157834; PIDN:ACC37201.1  
C:Genetics:  
A:Cross-references: FlyBase:FBgn0013119

Query Match	3.4%;	Score 257.5;	DB 2;	Length 1655;
Best Local Similarity	20.1%;	Pred. No. 8.2e-05;		
Matches 247;	Conservative 132;	Mismatches 489;	Indels 359;	Gaps 50;

[illegible]



```

QY 157 KSVYSILHWDHTEFYKNLPLKS-----IVNGGSWS 187
      : : : : :
Db 1396 LASFO-----GREVEYKLLLAYNANVEHRAKTGLTFLMECASGYDVGNLLIAAGADTN 1450
      : : : : :
QY 188 GEPPRRNSHTFNCRLMYKLPDPSEEGHD-----NOEHOKYETMO-CFA----- 231
      : : : : :
Db 1451 ASPEVOQKRD-----ALTIKAEKHEKVEMLNGDAVDVRNKKGTATLMLACNG 1502
      : : : : :
QY 232 -VEOPKSIKEEGEDLO-----SCLICVARVBMK-ERVPVLPSESEFTT RODLOGKITS 282
      : : : : :
Db 1503 YLSTAGALLKFGADPPMFDRKISPMMAARKHVEIVKTMVSAKOFPEODL----- 1556
      : : : : :
QY 283 LDSTYMAAMKPGMEDLVRC-----IQKTHAHEGESVYAKRHHEVLROGLAEFOIY 337
      : : : : :
Db 1557 -IRAQOETAETDIIKKCGECIDIIRSAKKAQASALAKLLELDEEKVQCEVK 1611
      : : : : :
QY 338 RESLSDGTLYAQTCKSLRSQTTNEPOLYISLHMLHREQNVCVMPDLTGQTMGKFLN 397
      : : : : :
Db 1612 KOKOKKIKKKEEK-KIKKQEAEPPE-----PEPEPEVPEP-EP 1651
      : : : : :
QY 398 ISSNSPAHQAICSGNPQDMLTSSNINFI--NGPEQMGMPRGREGSGGNHVSQMG 454
      : : : : :
Db 1652 VVISEPPEPV-----PIVEPEPEP--PKPR-----RNRKKTNP 1685
      : : : : :
QY 455 ATTPGGSNYALKMNPSSQSGMNPQPTSMLEPRHRMSPGAVGSPRIIPSPQSPACSLH 514
      : : : : :
Db 1686 DGVPKPKVYVEPKASTAEPESEMPYEPVIVTIP-----PPAKIH 1725
      : : : : :
QY 515 SPYGVCSSTGNHSHTNSSLNALQALS--EGHGVSLGSSLASBPKMLGNLQNSPVMPNP 571
      : : : : :
Db 1726 APWVSPGSESESEWCAGREKRVKSSKSGYAP--SSAGSQAQKESSTTSVSDQT 1783
      : : : : :
QY 572 PPLSKMGLSDSKOCFLGYGPESEGTQOASSCHPGEOKFTNPBNLPAYVSESRADQO-- 629
      : : : : :
Db 1784 PPT-----EITPRN-----ESSMKLTIPATAASRVIGKGG 1813
      : : : : :
QY 630 ---SRHDSKGQTKLLQLTTKSDQMEPSPLASLSLSDTNKDSFG----- 670
      : : : : :
Db 1814 SNNVAVREAGTALIEIKIQESNKQAEKVTLAKGTPEMVAYAMIIINMYIADADVLTA 1873
      : : : : :
QY 671 -----SLPGSGTHGSLKEKHILHRLLODSSPVDLAKLT-AEATGKDLSESS 720
      : : : : :
Db 1874 IRTVLKCNLSVASSFSSEGT--KSAVDSTYAPSSIPKSLSSASIJAROSASPIQOSS 1929
      : : : : :
QY 721 STAGS-----EVTIKOEPVSPKKENALLRYLDDKDTDIDG-----LP 760
      : : : : :
Db 1930 QRSKSHHOKDGGGCVWVHQRMAAREKEVEPLM-----ETKRISQSPKQAP 1976
      : : : : :
QY 761 ET--TPKLERLDSKTDPASTKLIAMKTEKEMSFEFGDQPGSELDNLELDDLQNSQL 818
      : : : : :
Db 1977 QIPSTQOOSKLOGRDQASET-LVRVAPAEVFAVPVASSIAAPSRNPVNLDRVIAPSL 2035
      : : : : :
QY 819 ---PQLEPDRPGAPAGSV-----DKQALINDLMQTAENSP-----VTPVGAOKTA 862
      : : : : :
Db 2036 RREPTTTLPLQVHPVOSVOSVOMQOOLAREPQKLAOPCLPPTIGQRYSQPISPOSS 2095
      : : : : :
QY 863 LRISQSTFNN-----PRGOLGRLLPNQNLPLDITLQSTGAGPEPPINSS 909
      : : : : :
Db 2096 VOVSQSSFSKAPGTBPSTDSRAPRPOQOQTQNNMTANELFDPLAGQCFRPTGMS 2155
      : : : : :
QY 910 PYSVIPIPGMGNGMGITGNQNLGNS-----STGMIGNSASRPTMP--G 952
      : : : : :
Db 2156 VAVTIVIPAKPSVNNQ--NNDKNGNSDDPDFSKRMFDEKGVGNIMWKSDSDSAMGLFS 2211
      : : : : :
QY 953 EMAPQSSAVAVTCAATTSAMNRPYOGGMINPASPIMRSSQPGQOQTQOQVMNIGPS 1012
      : : : : :
Db 2212 QFLPOLGA-----NSSLNNSNSKND--SNEGQNEFTSOLLIN----- 2248
      : : : : :
QY 1013 ELEMNMGPOYSQOQAPNPOTAPPE-----SLIPDQASFASQNRQPFSSPDLLCPH 1067
      : : : : :
Db 2249 -----SLIPNATSS-POGASTISSAPVQOPOSTSVTT--GLSLLELKGM 2290
      : : : : :

```

```

QY 1068 PAESPSDEGALLDQYLALRNFDEGLLEIDRALGIFELVSOQAVDPEQSSODSNIMLE 1127
      : : : : :
Db 2291 PASFAPSARDPNRSQPPLEFAR-----SQSNSV-----ANSTSTNIQOO 2328
      : : : : :
QY 1128 QKAPVFPQOYASQAOQMAQGSYSPMODPNFHTMGQRSYATLRNQRPRLGRLVGNQPN 1187
      : : : : :
Db 2329 QOQOIT--QHLDQOQALQOQOQRIQ--QFQOQYQOHQOSQ--SQPS 2368
      : : : : :
QY 1188 QLRLOLQHLRLOAQOONROPLMNQI-----SNVSN-----VNLTLRPG----- 1223
      : : : : :
Db 2369 DLMSKRFMSMQOQQOQOQOQLYQMQSILGQDQGSNLYNLHAAQLTHQESTGAPGFTSSQ 2428
      : : : : :
QY 1224 ---VPTQAPINAOMLAOPREILNOH-----LRQRMHQOQOYO 1259
      : : : : :
Db 2429 LANSYPSPEYTDASVILGQISMPSLSQRGIKQDFGNNDSDGVIAAILLEQOQOQKQGLA 2488
      : : : : :
QY 1260 QRTLMRGGCLNTPSPVAPSPGMPATMSNRIHQANAQOFPFPNNGISQPPRGFTGAT 1319
      : : : : :
Db 2489 QQSEFMHNSQ-----QOPFPGAPSNASANOSRLGMIQRPQPPR-VAPQAPGFSLSG 2540
      : : : : :
QY 1320 TPQSPIMSPMAHTQSPMMQO-----SQANPAYQAPEDINGMAQNGNSMFQSQSPPHF 1375
      : : : : :
Db 2541 NASTTNPSTSMQO--MQOQYQSSQOQPYGGMQAMQD-W-----NRLGQOQOQNAS 2589
      : : : : :
QY 1376 GQOQNTSMYSNNMNIN 1391
      : : : : :
Db 2590 GQOQHOSSSNKWSSN 2605
      : : : : :

```

## RESULT 37

```

T24158
hypothetical protein R11A8.7b - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
C:Accession: T24158; T24179
R:Bardill, S.
submitted to the EMBL Data Library, April 1996
A:Accession: T24158
A:Reference number: Z19846
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2584 <W11>
A:Cross-references: EMBL:270686; PIDN:CAB54289.1; GSPDB:GN00022; CESP:R11A8.7b
A:Experimental source: clone R10H10
R:Cummings, P.
submitted to the EMBL Data Library, March 1996
A:Reference number: Z19849
A:Accession: T24179
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2584 <M12>
A:Cross-references: EMBL:270310; PIDN:CAB54294.1; GSPDB:GN00022; CESP:R11A8.7b
A:Experimental source: clone R11A8
C:Genetics:
A:Gene: CESP:R11A8.7b
A:Map position: 4
A:Introns: 24/3; 80/3; 137/2; 177/1; 218/3; 348/3; 507/2; 709/3; 753/2; 924/1; 1020/3

```

```

Query Match 3.3%; Score 253.5; DB 2; Length 2584;
Best Local Similarity 18.2%; Pred. No. 0.00024;
Matches 272; Conservative 217; Mismatches 544; Indels 465; Gaps 63;

```

```

QY 100 SDVSS-TGGQVIDKDALGPMLEALDGFVNVLMLEGVYVSE--HYTQVLRVNOSELMN 156
      : : : : :
Db 1345 SEINSTRG---SKLGISPLMLASMG---HREATVILLEGSDINAOIETNRNTALT 1395
      : : : : :
QY 157 KSVYSILHWDHTEFYKNLPLKS-----IVNGGSWS 187
      : : : : :
Db 1396 LASFO-----GREVEYKLLLAYNANVEHRAKTGLTFLMECASGYDVGNLLIAAGADTN 1450
      : : : : :
QY 188 GEPPRRNSHTFNCRLMYKLPDPSEEGHD-----NOEHOKYETMO-CFA----- 231
      : : : : :

```

```

Db 1451 ASPVOQTDT-----ALFISAKEHEKFRMLNGDAADVANKKCTALMLACNG 1502
QY 232 -VSOPKSIKEGEDLO-----SCLICVARRVPMK -ERPVLPSSEPTTODLOGKITS 282
Db 1503 YLSTQALLEGADBPDMDNKIKSPMAAFKRGHVEIKYVNSAKOPPNEDL----- 1556
QY 283 LDTSTMRAMKPGMEDIVRC-----I0KFAH0EGESVSAYKRHHHEVLK0GLAFS0IY 337
Db 1557 -----IRAQ0TAETDDIKKKGCECIDLIRSAKK0AFSAELIAQ0LLELIDEBEK0K0VK 1611
QY 338 RPSLSDGLVA0QTSKILRSQ0TNEPOLVLSLHMLH0QVNCVNPDLTQ0TGMKPLNP 397
Db 1612 K0K0KDKKIKK0EEK-K1K0K0EA0EP0-----PP0EP0EP0EP0EP0EP0EP 1651
QY 398 ISSNSPAHQALCSGNPGQDMLSSNINPI---NGPK0GM0MPKRGFGSGGMNV0SQM 454
Db 1652 VVISEPVEPY-----PIVEEPK0EP---PKR-----RNKRKTNP 1685
QY 455 ATP0GSNTYALKMNSP0SS0PGMNP0PT0MLSPRHRMSPGVAGSPRIP0QSPAGSLH 514
Db 1686 DGVFKGPRVYVEPKASIAE0PSEMPY0PIVYTI-----PRAKIH 1725
QY 515 SPV0CVSGTSGSHSYTNSLAL0ALS---EGH0VSLGSSLASDPLKMGNI0NS0VNMNP 571
Db 1726 AP0VSPGYS0SESEMKAKGK0K0K0K0K0K0K0K0K0K0K0K0K0K0K0K0K0 1783
QY 572 PPLKMGSLD0K0CFGLYGE0P0EGT0GA0ESSCHP0EQ0K0ETND0NLP0AV0SERAD0GR 631
Db 1784 PPY-----EID0RN-----ESSKMLTIP0VAYASRIG--- 1810
QY 632 LHD0SG0TKL0LTLTTRSD0M0P0PLASLSLDTNKD0TSGSLPGSGSTHGTSLKEKHKILH 691
Db 1811 ---KGSNVNAV0EATGALIE---INKI0ESNK0A0ERTVLANGTPE--WVR0AMNITIN 1860
QY 692 RILD0S-----SSPYDLAKLTAEATGDL0S0ESSST-----A 723
Db 1861 YMITD0AVLVYDAIRTYL0RGM1SVASSP0S0EASIR0S0ASPIR00S0S0KSHH0KDS 1920
QY 724 PGSEVTIK0E0FVSPKKKENALRYLLDKD0TKD0G---LPEI---PKLERLDSKTDPA 776
Db 1921 GGGVNH0H0RMA0ARE0KEVEPL-----ETKRIS0SK0A0P0IPST000SKL0SD0QA 1972
QY 777 SNTKILAMKTEK0EMSEF0P0G0P0G0E0LDNLEI0LDL0NS0L---POL0P0TRGAP0AS 833
Db 1973 SET-LV0V0A0R0ENF0V0A0P0ASSI0A0P0SR0N0NVD0RY0A0P0L0RE0PTT0P0L0 2031
QY 834 V-----DK0AII0NDL0Q0LA0ENSP-----VTPVGA0Q0TALRI0S0FTNN----- 872
Db 2032 V0S0V0H0M0Q00L0AR0E0K0I0A0P0CLP0R0IG0RY0S0P0ISR0Q0S0V0S0S0F0K0 2091
QY 873 -----PRPG0LGRILP0N0NLEP0LTLT0SP0G0P0P0R0I0NS0P0SVI0P0P0GM0N0G0 927
Db 2092 D0S0AP0P0Q00T00N0MTT0AR0N0L0F0E0Q0L0F0K0P0T0G0M0SV0TVI0PAK0SVN0 2147
QY 928 N0G0MIGNS-----STGMIGNSAS0RPTMPS-----GEM0P0SA0AVI0CA0T0S 970
Db 2148 N0DK0NGNSD0F0FSKM0MF0DG0K0Y0N0MGK0S0D0S0ANG0L0F0S0L0P0L0GA----- 2197
QY 971 AM0R0V0G0G0M0IR0P0AS0IPMR0SS0P0C0R0T0L0S0V0N0IG0SELE0M0NG0F0Y0S0Q0 1030
Db 2198 ---NSL0NSNSK0NDS-----SNEM0C0N0E0I0S0LLI0-----SLP 2230
QY 1031 N0TP0R0PWE-----SLP0D0AF0S0N0R0P0R0GSS0P0DL0CH0PA0ES0P0SD0G0L0D0 1085
Db 2231 N0T0S0-P0G0A0ST0ISS0A0V00P0T0SS0VTT---GLS0L0E0K0G0M0P0AS0AD0P0R0S0P0 2286
QY 1086 AL0R0D0L0E0I0D0R0L0G0I0P0E0V0S0Q0A0V0P0E0F0S0D0S0N0I0E0K0A0P0F0 1145
Db 2287 FAR-----S0GSNSV-----AN0T0S0T0N0I0Q000000I---0H0L0Q000A0L0 2321
QY 1146 GS0Y0P0D0P0N0F0H0MG0R0P0S0AT0L0M0R0P0G0L0R0T0G0L0V0N0D0N0L0R0L0 1205
Db 2322 0000R0I0--0F000Y00H0S0S-----S0Q0P0S0D0MS0K0F0S0M0L0Q0S00000Q 2364

```

```

QY 1206 LMANOI-----SNVSN-----VNLTLRPG-----VPT0API0NA0MLA 1236
Db 2365 LY0M0Q0S0IG0D0G0GS0NLYNHLIA0QLLTH0ES0TAP0G0T0S0QLANSYIP0S0YTD0 2424
QY 1237 00RE0ELN0H-----L0R0M0H0Q00Y0Q0RTLM0RG0GLNMT0PSM 1277
Db 2425 0ISMP0LS0RG0IK0P0D0FN0NDS0GVIA0LLE0Q0Q00K0K0GLA0Q0S0F0M0NS0----- 2477
QY 1278 AP0SGM0P0ATM0NP0RIP0ANA0Q0F0P0P0NY0GIS0Q0P0D0P0GT0ATY0Q0S0PLM0 1337
Db 2478 0P0G0AP0S0NAS0S0R0LGM0I0P0R0P0P0F0-VAP0Q0AP0G0SS0L0GN0S0YT0N0P0 2534
QY 1338 M0Q0-----SQ0NP0AY0A0P0SD0ING0MA0GN0MG0S0M0SP0H0FG0A0NT0M0S0 1391
Db 2535 Y00T0G0SS0Q00PY0G0M0P0AMD-W-----NRLG0Q0Q0S0AS0G0Q0N0H0SS0 2583

RESULT 38
T32008
hypothetical protein K1066.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
R:Davidson, S.; Wohlmann, P.; Mullen, G.
submitted to the EMBL data library, July 1997
A:Description: The sequence of C. elegans cosmid K1066.
A:Reference number: J21111
A:Accession: J32008
A:Status: preliminary; translated from GB/EMBL/DDAJ
A:Molecule type: DNA
A:Residues: 1-1819 <DAV>
A:Cross-references: EMBL:AF016669; PIDN:AA066098.1; GSPDB:GND00020; CESP:K1066.3
A:Experimental source: strain Bristol N2; clone K1066
C:Genetics:
A:Gene: CESP:K1066.3
A:Map position: 2
A:Insertions: 41/2; 191/2; 556/2; 701/2; 1645/2; 1700/2; 1768/1; 1797/3

Query Match 3.3%; Score 252; DB 2; Length 1819;
Best Local Similarity 20.1%; Pred. No. 0.00018;
Matches 252; Conservative 133; Mismatches 482; Indels 384; Gaps 53;

QY 403 PAH0ALCSGN0G0D0MTLSSNINP0P0K0E0M0P0K0R0FG0SG0M0H0V0SG0M0ATT0P0GSN 462
Db 56 P0H0H0Q0A0Q0Y0G0-ATG0STN-----GGG0Q0Q0M---TSMY0G0NDY0D0H0LH0N0Q0H0A0 106
QY 463 YALKMNSP0S0SP0M--NP0Q0PT-SM0SPR0HRMSP0V0A0SPRIP0P0S0F0S0AGSLH0SPV 519
Db 107 ST0Q0F0H0P0R0P0P0Y0D0P0SS0T0SS0SLP0LHYA--Y0L0R0P0P0S0N0R0T0P0Q0L0Y0PV 164
QY 520 CS0T0GNSH0YTNS0LNA0LSE0H0GVSLGSSLAS0PLKMGNI0NS0VNMNP0PLSKMGS 579
Db 165 VE0AG0Q0AY0A0---QY0Q0A0Q0S0M0R0SGA0GVNSA-----L0Q0K0L-----P0LS0IT 207
QY 580 LD0K0C0F0GL0G0PS---EGT0G0-----A0ESS0CH0P0E-----Q0KET0D0N0L 617
Db 208 ISS0SA0GSS0IS0AP0T0S0P0T0S0LTLT0P0ST0SS0M0P0R0K0T0P0N0A0SS0SLIK0 267
QY 618 PPAV0SER0A0G0S0LH0S0K0G0TKLQ--LTLT0KSD0M0P----- 654
Db 268 Q0R0Y0D0F0E0Y0A0RV0S0IM-SK0N0K0L0M0E0P0LTL0L0P0LA0P0L0P0PK0S0Y0Q0 326
QY 655 -----SPLAS0LSDP0NK0D0ST0GSL- 672
Db 327 ANA0R0L0R0H0T0G0SA0HA0P0LA0M0L0R0S0P0P0CA0S0P0V0A0P0R0T0S0Q0HN0S0 386
QY 673 ---PG0SGT0HT0GSL---KE0KH0KIL-----HRL0D0SS0PVD0LAKIL0E0AT0K0D0L0 722
Db 387 LSY0SS0SS0R0S0SLY0SP0LH0D0LVC0NP0V0M0L0SG0Y0E0K0DD0P0MLY0G0P0S0----- 440
QY 723 AP0GSEV0TIK0E0PVS0PKKENA0LRYLLDKD0T0D0IG0LPE-ITPKLERLDSK0T0P0AS0N 781

```

```

Db      441 IWSRDSFHSFPPSASH-----DQLMDHLGFPLDPLHLHLS-FDSADHRKE 489
OY      782 IAKKTEK--EEMSFEP-GDOPGE-----LDNLEELID---LONSOLQLPPDTR-PC 828
Db      490 TPKECHERDELMTLDPTRPPCGSEFRGIGIDMDPLSLDDEDEPLMSSESLSSSGRN 549
OY      829 APAGSVDKQAIINDLMQTLAENSPT-----PVGAOKTALRISQSTFNNPRGQLG 879
Db      550 TPAVAFRRP-----TKRKRPATFRPSKKEASEGPEKNSAIIAALRKEPAAPQ 601
OY      880 RLHPNMLPLDITLQSTGAGPR-----PIRNSPY 912
Db      602 QLOFQOYQPSQFQAPYGGSLPSISASVLSASTSAAABERSEMFTSPITVSAHP 661
OY      913 VI-----PQPMGMCNOCMIGNOCNIGNSGTMTIGNSRPTMGEMAPQS 958
Db      662 YIHQDHPHQOQOQSSPLEDLNDESDADDSDSSSTVSNSTTTTATTTSKST 721
OY      959 SAVRYTC-----AATTSAMNR-----PVGGMIRNPASIPMRPSQPGQOTLOSVM 1007
Db      722 GNPLFTEHCARQLCSMSNLKRRATCKVAASSNSAAS--RPSQSTPATAPAT-- 776
OY      1008 NIGPSELEMMGGRQYSQOAPRNOTAPRPESILPIDQAFASQNR-----QPRGSS 1059
Db      777 -----PMLQASQAPQLQAP-POS--PMETATVYTKTVPPSVANTMTE 820
OY      1060 PDLCLPFAPESPSDGALLDQULYLRNFDGLEIDRALGIPELYSQOAVDPEQSS 1119
Db      821 KAOLISKPRSSQITFSASSMTVGDLRAQOQKQDQIOT-----QFOQ 867
OY      1120 QDSNIMLEOKARVFRQOYASQAOAGSYSPMODPNEHTMGORSPATYLRMPRRGLRPT 1179
Db      868 QOQ-----QRFQHQQOQOQAG-----RIPRRPRL 894
OY      1180 GLVONOPNQLRLQNLHQAQONRPLMNQ-----ISVSNVNLTLRPGVTPQAP 1230
Db      895 NOVONPRQOVQONHONQONMLNPIRQLQSPRRPRKGLIEHKNDLVL---ITSEPL 950
OY      1231 NAOMLAQOREILNQLRQRMHQOQOQVQOQRTLMRGOGL-NMTPSAVAFSGMATWSNP 1289
Db      951 AERRDKARR-----SSEGLVAVTSPLPPIQLPQRQAP 984
OY      1290 RTRQANAQOFRPRPNCISQORPRGTGATTPQ---SPLMSPRMATQSPMAQOQANPA 1346
Db      985 -AASRQOQOQOP-RVAYUVQ-----FNGRPLRPMQRLPRLPHNOQOQOHLMDSQNN-- 1034
OY      1347 YQAPSDI-----NGAOGNMGMSMFSSQSPRRHGGQANTSMYSNNMINVSMATNTGMS 1402
Db      1035 YQOYQOYQOVQOQONLQNHQOQHHQOQOQAPRGKRSKSHNV-----GKME 1086
OY      1403 SMOQMTQISMTSVTSYSTS-----GLSSMGREQVNDPALR 1438
Db      1087 QEAQRQSSPLDITTSVRLSIVENHNHIMKGRPLEQGOSSVDSQSTAPRSR 1137

RESULT 39
S39161
CRBB-binding protein - mouse
C:Species: Mus musculus (house mouse)
C:Date: 25-Feb-1994 #sequence_revision 10-Nov-1995 #text_change 20-Apr-2000
C:Accession: S39161
R:Chitvia, J.C.; Kwok, R.P.S.; Lamb, N.; Hagiwara, M.; Montminy, M.R.; Goodman, R.H.
Nature 365, 855-859, 1993
A:Title: Phosphorylated CRBB binds specifically to the nuclear protein CBP.
A:Reference number: S39161, PMID:94019866
A:Accession: S39161
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-2441 <CHR>
* C:Superfamily: unassigned bromodomain proteins; bromodomain homology
F:1112-1109/Domain: bromodomain homology <BRO>

```

```

Query Match      3.3%; Score 252; DB 2; Length 2441;
Best local similarity 20.6%; Pred. No. 0.00026;
Matches 248; Conservative 142; Mismatches 442; Indels 372; Gaps 61;

OY      392 GKPLNP-----ISSNPAHQALCSGNPGQMTLSSNINPFINPGKMGMPKGRFGSGGM 447
Db      78  GSSINPGIGVNASSPVQOGL--GGOAQOQPNSTN-----MMSLGAM-GK 119
OY      448 NHVSGMAQTTPQGSNVALKMNPSQSSPGMNPQOPRSM-TSPHRMRSPGVASPRIPRQ 506
Db      120 SPLNGDSSTP-----NLPRQAISTGPRPPASQALNPQAOQVLYNS----- 163
OY      507 FSPAGLSHPVGVCSSTGNSHSY-----TNSLNALQALSEHGVSAGSLASPLDKMN 561
Db      164 -SPATISQTP-GICMANANQTHPHGLNSISGHSLNQAOQOQAOYVMSNL----- 212
OY      562 LQNSPVNMNRPPLSKMSLSDKDFGLYGPSEGTGQAESSCHPEQXETDNPILPRAV 621
Db      213 -----GAAGRGAGAMPYPA-----PAMQAT 234
OY      622 SSEPAD-----GQSRHDSK--GOTKILLTLTKSDMEPSPLASSISDTKKDST 669
Db      235 SSVLAETTLQVSPOMAGHGLNTPAQAGMTKMGMTGT-----SPGQPSQTGGQM 287
OY      670 GSLPGSGSTHGTSLKEKHILHRLQDSSPYDLAKTLTAATGKDLQSESSTAPESEYT 729
Db      288 GA-----TGYNPLQASKQSVNSL---PAFPDI-KNTSVTVYPPNNSQLOTSGVTPQA 338
OY      730 IKQEPVSPKKENALLRYLQDKDQDKDGLPEITPKLERLDSKTDPASTKLIAMKTEKE 789
Db      339 IATGPADPEKRLIQOQLV-----LLLHAKHCOQR 369
OY      790 EMSFERGDDGSELDN---LEETIDDLQNSQLPQLPPTRRPAGASVDKQAIINDLMQL 846
Db      370 EQA--NGEVACSLPRCTMKVNLNHTHQAP-----KACVNAACASSRQIISHWKNC 421
OY      847 TAENSPVTPVGAOKTALRISQSTFNNPRRQGLRLLPNQNLPLDITLQSPGTGAPRRPJR 906
Db      422 TRHDCPY-----CLPKNASDKRNOQTILG----- 446
OY      907 NSSPYVITPQPMGNOGMLGNGNLGNSSTGIGASASRPTMPSERMAPQSSAVRYTCA 966
Db      447 --SPAS-----GIQNTIGVGAQOQNAVTSL--SNPNPIDP-----SSMORAUYA 486
OY      967 AATSAMNRPVQGGMIRNPASIPMRPSQPGQROT-LQSOVMN-IGPSELEMMGGRQYS 1024
Db      487 LGRLPNQF-----QTOLOPQVPGQORAPRPAHQOKRTILNALNNPMSVPAAGITTD 538
OY      1025 QOQAPRNOTAPRPESILPIDQAFASQONRQF--GSSPDDL-LCPHRAESPSDEGA- 1078
Db      539 QQ--PMLIS--ESALP--TSIGATN-PLMNDGSSNGSITSLSTIPTAAPRSSTGYR 588
OY      1079 -----LDDQLYAL-RNFDGLEIDRALGIPELYSQOAVDPEQSSODSN 1123
Db      589 KGNHEHVTDLRSHLVAKLVQALFPTPRDALKDRRR--ENLVAIVAKKVEGDYIESANSR 646
OY      1124 ----IMLEOKARVFRQOYASQAOAGSYSPMODPNEHTMGORSPATYLRMPRRGLRPT 1179
Db      647 DEYVHLLAEKI-----YKQKLEEKRRTRRLHKQGI--LGNDPALPASAQAP-PVIRPA 697
OY      1180 GLVONOPNQLRLQNLHQAQONRPLMNQISNVSNVNLTLR----- 1221
Db      698 QSVRRPGRPLRLPV-NRMQVSQ-----MNSFNPNLSIGNVQLPQAPRGPRASPMNHSVQM 752
OY      1222 -----PGV---PTQAPINAQMLAQOREILNQLRQRMHQOQOQOQRTLMNR-----GO 1268
Db      753 NSMASVPGMAISPSRMPQPRNMGTANNIMAPQONQFLPQONQPPSSGASVNSVGM 812
OY      1269 GLMWTSPNVAPSGMP-ATMSNP---RLPQANAQOFRPPRYVSGSQOP----- 1311
Db      813 GQPAQAGVYSGQGPQALPPLNMLAPQAS--QLPCPP---VTSQPLHPPTPPASTAAG 867

```





**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 7, 2002, 10:38:29 ; Search time 16.67 Seconds  
(without alignments) 3400.445 Million cell updates/sec

Title: US-09-842-256-2

Perfect score: 7631  
Sequence: 1 MSGMGENTSDPSRAETRRK.....NQLPGMDIKQEGDTTRK 1464

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listed first 45 summaries

Database: SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7626	99.9	1464	1 NCO2_HUMAN	Q15596 homo sapien
2	7216	94.6	1462	2 NCO2_MOUSE	Q61026 mus musculu
3	7160.5	93.8	1465	1 NCO2_RAT	Q9WU19 rattus norv
4	5276.5	69.1	1516	1 NCO2_XENLA	Q9W705 xenopus lae
5	328	4.3	1507	1 STMA_DROME	Q24167 drosophila
6	317	4.2	846	1 CLOC_HUMAN	Q15516 homo sapien
7	310.5	4.1	644	1 ARNT_DROME	Q15945 drosophila
8	309.5	4.1	855	1 CLOC_MOUSE	Q08785 mus musculu
9	289	3.8	816	1 NP2A_MOUSE	P97460 mus musculu
10	289	3.8	1531	1 NP2A_HUMAN	Q94916 homo sapien
11	284.5	3.7	1902	1 SMF1_HUMAN	Q14497 homo sapien
12	284	3.7	874	1 PAS1_MOUSE	P97481 mus musculu
13	278.5	3.6	870	1 PAS1_HUMAN	Q09814 homo sapien
14	277	3.6	2414	1 P300_HUMAN	Q09472 homo sapien
15	273.5	3.6	766	1 SIM1_HUMAN	P81133 homo sapien
16	270	3.5	1589	1 PHF_DROME	P39769 drosophila
17	268.5	3.5	826	1 HTRF_HUMAN	Q16665 homo sapien
18	268.5	3.5	3726	1 ABL1_MOUSE	Q61329 mus musculu
19	267	3.5	583	1 BMAL1_HUMAN	Q00327 homo sapien
20	266.5	3.5	2090	1 N214_HUMAN	P35658 homo sapien
21	266	3.5	765	1 SIM1_MOUSE	Q61045 mus musculu
22	266	3.5	1211	1 BUN2_DROME	Q24523 drosophila
23	262	3.4	706	1 ARN2_HUMAN	Q9H222 homo sapien
24	262	3.4	822	1 HTRF_MOUSE	Q61221 mus musculu
25	260.5	3.4	712	1 ARN2_MOUSE	Q61324 mus musculu
26	259	3.4	824	1 NP2A_HUMAN	Q09743 homo sapien
27	258.5	3.4	3969	1 HRX_HUMAN	Q03164 homo sapien
28	258	3.4	2004	1 MOZ_HUMAN	Q92794 homo sapien
29	252	3.3	2441	1 CBP_MOUSE	P45481 mus musculu
30	251	3.3	1794	1 YAVI_SCHPO	Q10172 schizosacch
31	250	3.3	1596	1 MAM_DROME	P21519 drosophila
32	249.5	3.3	3703	1 ABL1_HUMAN	P15911 homo sapien
33	249	3.3	1023	1 CLOC_DROME	Q61735 drosophila

34	248	3.2	790	1 ARNT_RABIT	Q02748 oryctolagus
35	248	3.2	2175	1 HMCU_DROME	P10180 drosophila
36	248	3.2	3828	1 TRX_DROME	Q24742 drosophila
37	243.5	3.2	958	1 TRH_DROME	Q24119 drosophila
38	243	3.2	2442	1 CBP_HUMAN	Q92793 homo sapien
39	243	3.2	3866	1 HRX_MOUSE	P55200 mus musculu
40	241.5	3.2	3726	1 TRX_DROME	P20659 drosophila
41	241	3.2	1603	1 PSC_DROME	P35820 drosophila
42	239	3.1	673	1 SIM1_DROME	Q05709 drosophila
43	237	3.1	413	1 CYCL_DROME	Q61734 drosophila
44	236	3.1	1403	1 PRO_DROME	P29617 drosophila
45	234.5	3.1	1210	1 AF4_HUMAN	P51825 homo sapien

## ALIGNMENTS

RESULT	ID	NC02_HUMAN	STANDARD	PRT	1464 AA.
AC	Q15596				
DT	16-OCT-2001	(Rel. 40, Created)			
DT	16-OCT-2001	(Rel. 40, Last sequence update)			
DT	16-OCT-2001	(Rel. 40, Last annotation update)			
DE	Nuclear receptor coactivator 2 (transcriptional intermediary factor 2).				
DE	NCO2 OR TIF2.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISUE=Placenta;				
RX	MEDLINE=96312964; PubMed=8670870;				
RA	Voegel J.J., Helne M.J.S., Zechel C., Chambon P., Gronemeyer H.;				
RT	"TIF2, a 160 kDa transcriptional mediator for the ligand-dependent activation function AF-2 of nuclear receptors.";				
RL	EMBO J. 15:3667-3675(1996).				
RN	[2]				
RP	CHARACTERIZATION.				
RX	MEDLINE=98094372; PubMed=9430642;				
RA	Voegel J.J., Helne M.J.S., Tini M., Vivat V., Chambon P., Gronemeyer H.;				
RT	"The coactivator TIF2 contains three nuclear receptor-binding motifs and mediates transactivation through CBP, binding-dependent and -independent pathways.";				
RL	EMBO J. 17:507-519(1998).				
CC	-1- FUNCTION: TRANSCRIPTIONAL COACTIVATOR FOR STEROID RECEPTORS AND NUCLEAR RECEPTORS. COACTIVATOR OF THE STEROID BINDING DOMAIN (AF-1).				
CC	(AF-2) BUT NOT OF THE MODULATING N-TERMINAL DOMAIN (AF-1).				
CC	-1- SUBCELLULAR LOCATION: Nuclear.				
CC	-1- SIMILARITY: CONTAINS 1 PAS (PER-ARNT-SIM) DIMERIZATION DOMAIN.				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to <a href="mailto:license@sib-sib.ch">license@sib-sib.ch</a> ).				
CC	EMBL: X97674; CAA66263.1; -				
DR	TRANSPAC: T02483; -				
DR	MIM: 601993; -				
DR	InterPro: IPR001974; GelsoLin.				
DR	InterPro: IPR001092; HLH_dlm.				
DR	InterPro: IPR001610; PAC.				
DR	InterPro: IPR000014; PAS.				
DR	Pfam: PF00626; GelsoLin; 1.				
DR	Pfam: PF00010; HLH; 1.				
DR	Pfam: PF00989; PAS; 1.				
DR	SMART: SM00353; HLH; 1.				

DR SMART: SM00086; PAC: 1.  
 DR SMART: SM00091; PAC: 1.  
 DR PROSITE: PS0112; PAC: 1.  
 KW Transcription regulation; Activator; Nuclear protein.  
 FT DOMAIN 119 183 PAS.  
 FT DOMAIN 1254 1260 POLY-GLN.  
 SQ SEQUENCE 1464 AA; 159156 MW; 0A61AA5D1878304B CRC64;

Query Match 99.9%; Score 7626; DB 1; Length 1464;  
 Best Local Similarity 99.9%; Pred. No. 0;  
 Matches 1463; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

QY 1 MSGGENTSPRAETKREKRECPQGLSPKRNTERKRNREDEKNTIELAELEIPIANNDI 60
DB 1 MSGGENTSPRAETKREKRECPQGLSPKRNTERKRNREDEKNTIELAELEIPIANNDI 60
QY 61 DNFNFKPKCAILKETKQJROIKEDEKAAANIDEVOKSDVSTGGGVTDKDALGPMML 120
DB 61 DNFNFKPKCAILKETKQJROIKEDEKAAANIDEVOKSDVSTGGGVTDKDALGPMML 120
QY 121 EALDGEFFVNLGNVNVFSENVTOYLRYNOEELMKNKSVYSLHVGDTHEFVKNLIPKSI 180
DB 121 EALDGEFFVNLGNVNVFSENVTOYLRYNOEELMKNKSVYSLHVGDTHEFVKNLIPKSI 180
QY 181 VNGSGSGEPRRNSHTFNCRMLVPLPDESEGHNOEAKOXYETMOCAVNSOPKSIKE 240
DB 181 VNGSGSGEPRRNSHTFNCRMLVPLPDESEGHNOEAKOXYETMOCAVNSOPKSIKE 240
QY 241 EGEDLOSLICVARVPMKREPVLPSESEFTTRQDLOGKITSLDTSTRAMKPGMEDLV 300
DB 241 EGEDLOSLICVARVPMKREPVLPSESEFTTRQDLOGKITSLDTSTRAMKPGMEDLV 300
QY 301 RRCIOFHAOHEBESVYAKRHHEVLRQGLAFSQTIRFSLSDGTLVAOATKSLRSQT 360
DB 301 RRCIOFHAOHEBESVYAKRHHEVLRQGLAFSQTIRFSLSDGTLVAOATKSLRSQT 360
QY 361 TNEPOLVLSLHMLHREONVCMNPDLTGOMGKPLNPISNSPAHQALCGNQGDMTSL 420
DB 361 TNEPOLVLSLHMLHREONVCMNPDLTGOMGKPLNPISNSPAHQALCGNQGDMTSL 420
QY 421 SNINPENGKQMGPMGRFGSGGMNHVSGMATTPOGSNALKMNSPQSGMNPNG 480
DB 421 SNINPENGKQMGPMGRFGSGGMNHVSGMATTPOGSNALKMNSPQSGMNPNG 480
QY 481 OPTSMLSPRHRMSPVAGSPRIIPSPQSPAGSLHSPVGVCSSTGNSHSTYNSLALQAL 540
DB 481 OPTSMLSPRHRMSPVAGSPRIIPSPQSPAGSLHSPVGVCSSTGNSHSTYNSLALQAL 540
QY 541 SEGHGVLGSSLASPLDKMGNLONSPPVNNPPLSKMGLSDSKDCGIVGEPEEGTGA 600
DB 541 SEGHGVLGSSLASPLDKMGNLONSPPVNNPPLSKMGLSDSKDCGIVGEPEEGTGA 600
QY 601 ESSCHPGEKETNDPMLPRAVSSERADGQSLHDSKQTLQTLTTKSDOMPSPGLASS 660
DB 601 ESSCHPGEKETNDPMLPRAVSSERADGQSLHDSKQTLQTLTTKSDOMPSPGLASS 660
QY 661 LSDTNDSTGSLPGSGSTGTSTLKEKHKLHRLDQSSSPVDLAKTAAATGKDLQESS 720
DB 661 LSDTNDSTGSLPGSGSTGTSTLKEKHKLHRLDQSSSPVDLAKTAAATGKDLQESS 720
QY 721 STAPGSEVITKOEVPSPKKENALLRYLDKDDTKDIGPETTPKLERLDSKTDPSNNK 780
DB 721 STAPGSEVITKOEVPSPKKENALLRYLDKDDTKDIGPETTPKLERLDSKTDPSNNK 780
QY 781 LIAMTKEKMSFEPDQSGSELDNLEILLDLQNSQLPQLPDRPGAPAGSVDAQAT 840
DB 781 LIAMTKEKMSFEPDQSGSELDNLEILLDLQNSQLPQLPDRPGAPAGSVDAQAT 840
QY 841 NDIMOLTEKNSVTPVGAQKTAIRISQSFNNPRGQGLRLPNONLPLDITLQSPGTAG 900
DB 841 NDIMOLTEKNSVTPVGAQKTAIRISQSFNNPRGQGLRLPNONLPLDITLQSPGTAG 900

```

```

QY 901 PEPPIRNSPYSVPIPOGMNGNOGMIGNOGLNGSSTGMIGNSASRPTMPSEGAPOSSA 960
DB 901 PEPPIRNSPYSVPIPOGMNGNOGMIGNOGLNGSSTGMIGNSASRPTMPSEGAPOSSA 960
QY 961 VRVTCATTSAMNRPVGGMRNPAASIPMPSSOPQOROTLOSQVANNIGSELEMMNG 1020
DB 961 VRVTCATTSAMNRPVGGMRNPAASIPMPSSOPQOROTLOSQVANNIGSELEMMNG 1020
QY 1021 POVSQOQAPPNOTAPWPESTILPIDQASFASQNRQPFSSPDLLCPHAAESPDEGALL 1080
DB 1021 POVSQOQAPPNOTAPWPESTILPIDQASFASQNRQPFSSPDLLCPHAAESPDEGALL 1080
QY 1081 DQYLALRNFDGELEIDRALGIPELVSOQAVDEQFSSODSNIMLEOKAPVPEQYASQ 1140
DB 1081 DQYLALRNFDGELEIDRALGIPELVSOQAVDEQFSSODSNIMLEOKAPVPEQYASQ 1140
QY 1141 AQMAQGSYSPMODPNEFTMGQRPSTAVLRMQPRGLRPTGLVONQPOLRLQDHRLOAQ 1200
DB 1141 AQMAQGSYSPMODPNEFTMGQRPSTAVLRMQPRGLRPTGLVONQPOLRLQDHRLOAQ 1200
QY 1201 QNRQPLMNOISNVSNVLTLPQVPTQAPINAQMLAQROREILNOLROROMHQOQVOQ 1260
DB 1201 QNRQPLMNOISNVSNVLTLPQVPTQAPINAQMLAQROREILNOLROROMHQOQVOQ 1260
QY 1261 RTLMRQGLNMTPSVYAPSGMPATMSNPRIPOANAQOFPFPNYGISQPDGFTGATT 1320
DB 1261 RTLMRQGLNMTPSVYAPSGMPATMSNPRIPOANAQOFPFPNYGISQPDGFTGATT 1320
QY 1321 POSPLMSPRHAHTQSPMAQSOANPAVQAPSDINGMAQGMGNSMFSQDSPHFGQAN 1380
DB 1321 POSPLMSPRHAHTQSPMAQSOANPAVQAPSDINGMAQGMGNSMFSQDSPHFGQAN 1380
QY 1381 TSMYSNNMNIIVSMATNTGMSMNOQTQISMTSVTSVTSGLSSMGPCQVNDPALRG 1440
DB 1381 TSMYSNNMNIIVSMATNTGMSMNOQTQISMTSVTSVTSGLSSMGPCQVNDPALRG 1440
QY 1441 NLEPNQLPGMDMTKQEGDITTRKYC 1464
DB 1441 NLEPNQLPGMDMTKQEGDITTRKYC 1464

```

RESULT 2  
 NCO2\_MOUSE  
 ID NCO2\_MOUSE STANDARD; PRT: 1462 AA.  
 AC Q61026; P97759; O09001;  
 DT 16-OCT-2001 (Rel. 40; Created)  
 DT 16-OCT-2001 (Rel. 40; Last sequence update)  
 DT 16-OCT-2001 (Rel. 40; Last annotation update)  
 DE Nuclear receptor coactivator 2 (transcriptional intermediary factor 2)  
 DE (Glucocorticoid receptor-interacting protein 1) (GRIP-1).  
 GN NCOA2 OR TIF2 OR GRIP1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sclirognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ICR; TISSUE=Brain;  
 RA MEDLINE=97265407; PubMed=9111344;  
 RA Hong H., Kohli K., Garabedian M.J., Stallcup M.R.;  
 RT "GRIP1, a transcriptional coactivator for the AF-2 transactivation  
 domain of steroid, thyroid, retinoid, and vitamin D receptors.";  
 RL Mol. Cell. Biol. 17:2735-2744(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=97336097; PubMed=9192892;  
 RA Torchia J., Rose D.W., Inostroza J., Kamei Y., Westin S., Glass C.K.,  
 RA Rosenfeld M.G.;  
 RT "The transcriptional co-activator p/CIP binds CBP and mediates  
 RT nuclear-receptor function.";  
 RL Nature 387:677-684(1997).  
 RN [3]  
 RP SEQUENCE OF 322-1119 FROM N.A.

CC STRAIN-ICR; TISSUE=Embryo;  
 RA MEDLINE=96209838; PubMed=8643509;  
 RX Hong H., Kohli K., Trivedi A., Johnson D.L., Stallcup M.R.;  
 RT "GRIP1, a novel mouse protein that serves as a transcriptional  
 RT coactivator in yeast for the hormone binding domains of steroid  
 RT receptors.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 93:4948-4952(1996) **★**  
 CC -1- FUNCTION: TRANSCRIPTIONAL COACTIVATOR FOR STEROID RECEPTORS AND  
 CC NUCLEAR RECEPTORS. COACTIVATOR OF THE STEROID BINDING DOMAIN  
 CC (AF-2) BUT NOT OF THE MODULATING N-TERMINAL DOMAIN (AF-1).  
 CC -1- SUBCELLULAR LOCATION: Nuclear.  
 CC -1- TISSUE SPECIFICITY: UBIQUITOUS.  
 CC -1- SIMILARITY: CONTAINS 1 PAS (PER-ARNT-SIM) DIMERIZATION DOMAIN.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: U39060; AAC53151.1; -;  
 DR EMBL: AF000582; AAB61575.1; -;  
 DR TRANSFAC: T02482; -;  
 DR MGD: MGI:1276533; Ncoaz.  
 DR InterPro: IPR001974; Gelsolin.  
 DR InterPro: IPR001092; HLH\_dim.  
 DR InterPro: IPR001610; PAC.  
 DR InterPro: IPR000014; PAS.  
 DR Pfam: PF00626; Gelsolin; 1.  
 DR Pfam: PF00010; HLH; 1.  
 DR Pfam: PF00989; PAS; 1.  
 DR SMART: SM00353; HLH; 1.  
 DR SMART: SM00086; PAC; 1.  
 DR SMART: SM00091; PAS; 1.  
 DR PROSITE: PS50112; PAS; 1.  
 DR KW Transcription regulation; Activator; Nuclear protein.  
 FT DOMAIN 119 183  
 FT CONFLICT 51 51 D -> E (IN REF. 2).  
 FT CONFLICT 140 141 SE -> FR (IN REF. 2).  
 FT CONFLICT 194 194 T -> S (IN REF. 2).  
 FT CONFLICT 251 320 CVARVPKMERPTLPSSSEFTTRDLOGKITSLDTSTWRAA  
 FT MKPGEDELVRCIOKFTFHOGESLSYAK -> VMHEDPHE  
 FT GKTNSLIRKLHPGPRRODHFTHGHESRHEHSGRSG  
 FT KDAFRSTSHMKSGSLYMPR (IN REF. 2).  
 FT G -> S (IN REF. 2).  
 FT S -> N (IN REF. 2).  
 FT E -> K (IN REF. 2).  
 FT EE -> KK (IN REF. 2).  
 FT R -> C (IN REF. 2).  
 FT T -> S (IN REF. 2).  
 FT N -> Y (IN REF. 2).  
 FT PAVRTCAATGAMNRPVQGMIR -> TSCSEPLVLLPLV  
 FT PRDQSKEARFG (IN REF. 2).  
 FT R -> G (IN REF. 2).  
 FT P -> L (IN REF. 2).  
 FT C -> G (IN REF. 2).  
 FT P -> L (IN REF. 2).  
 FT CONFLICT 959 982  
 FT CONFLICT 991 991  
 FT CONFLICT 996 996  
 FT CONFLICT 1407 1407  
 FT CONFLICT 1446 1446  
 FT SEQUENCE 1462 AA: 158511 MW: ACAD18979FCDCAB5 CRC64;  
 Query Match 94.6%; Score 7216; DB 1; Length 1462;  
 Best Local Similarity 94.1%; Pred. No. 0;  
 Matches 1377; Conservative 43; Mismatches 42; Indels 2; Gaps 1;

QY 121 EALDGFEEVNLGNVFESENVTOYLRYNOBELMKNKSYSTLHYGDHTEFYKNLLPKSI 180  
 DB 121 EALDGFEEVNLGNVFESENVTOYLRYNOBELMKNKSYSTLHYGDHTEFYKNLLPKSI 180  
 QY 181 VNGSGNSGEPARRNSHTFNCRLIVPLPDSEEGHNDNOEAHQYETMQCAFAVSQPSKIKE 240  
 DB 181 VNGSGNSGEPARRNSHTFNCRLIVPLPDSEEGHNDNOEAHQYETMQCAFAVSQPSKIKE 240  
 QY 241 BEEDLQSLICVARRVPKMERPVLPSSSEFTTRDLOGKITSLDTSTWRAAMPQMEDLY 300  
 DB 241 BEEDLQSLICVARRVPKMERPVLPSSSEFTTRDLOGKITSLDTSTWRAAMPQMEDLY 300  
 QY 301 RRCIOKFAHQHEGESVAKRHHHEVLRGLAFSQTIRPSTLSDGLTYAAQTSKILRSQT 360  
 DB 301 RRCIOKFAHQHEGESVAKRHHHEVLRGLAFSQTIRPSTLSDGLTYAAQTSKILRSQT 360  
 QY 361 TNEPOLVLSLHMLHREOVNVCVNPDLTGQTMGKPLNPISNSPAHQALCSGNPGODMTLS 420  
 DB 361 TNEPOLVLSLHMLHREOVNVCVNPDLTGQTMGKPLNPISNSPAHQALCSGNPGODMTLS 420  
 QY 421 SNINFPINGPKQWGMKRGFGSGGMNHVSGMQATTPQGSNYALKMNSPQSSPGMNP 480  
 DB 421 SNINFPINGPKQWGMKRGFGSGGMNHVSGMQATTPQGSNYALKMNSPQSSPGMNP 480  
 QY 481 QPTSLSPRHRSKPVAGSPRIIPSOFPAGSLHSIPVGCSTGNSHTNSLNAALQAL 540  
 DB 481 QPTSLSPRHRSKPVAGSPRIIPSOFPAGSLHSIPVGCSTGNSHTNSLNAALQAL 540  
 QY 541 SEGHEVSLGSSLASPDLMKGNLQNSPVNMPPLPKMGSIDSKDCDFGLYGESEEGTGOA 600  
 DB 541 SEGHEVSLGSSLASPDLMKGNLQNSPVNMPPLPKMGSIDSKDCDFGLYGESEEGTGOA 600  
 QY 601 ESSCHPEGEKFTNDNPLPVAASERADQSRUHSKSGOTKLLQLLTYSKSDONEPSPASS 660  
 DB 601 ESSCHPEGEKFTNDNPLPVAASERADQSRUHSKSGOTKLLQLLTYSKSDONEPSPASS 660  
 QY 661 LSDTKKDSGSLPGSGSTHGSLTKKHKHLHLLDSSSPVDLAKTLAEATKELDSOSS 720  
 DB 661 LSDTKKDSGSLPGSGSTHGSLTKKHKHLHLLDSSSPVDLAKTLAEATKELDSOSS 720  
 QY 721 STAPGSEYTIKQEPVSPKKENALRYLLDKDDTDIGLPEITPKLERLSDKTDPAWTK 780  
 DB 721 STAPGSEYTIKQEPVSPKKENALRYLLDKDDTDIGLPEITPKLERLSDKTDPAWTK 780  
 QY 781 LIAMTKTEKMSFEFGDDPGSELNLELDDLDQNSQLPQLFPDTRPGAPAGSVDKAII 840  
 DB 781 LIAMTKTEKMSFEFGDDPGSELNLELDDLDQNSQLPQLFPDTRPGAPAGSVDKAII 840  
 QY 841 NDLMQUTAEKSPVTVGAOKTALRISQSTFNNPRRGOLGRLLPNONLPLDITLOSPTGAG 900  
 DB 841 NDLMQUTAEKSPVTVGAOKTALRISQSTFNNPRRGOLGRLLPNONLPLDITLOSPTGAG 900  
 QY 901 PPPTRNSSPSVIIPQPMGNGMIGNGLNLSSTMGINSSTMGINSASPTMPSGEMAPQSSA 960  
 DB 901 PPPTRNSSPSVIIPQPMGNGMIGNGLNLSSTMGINSSTMGINSASPTMPSGEMAPQSSA 960  
 QY 961 VRYTCAATTSAAMNRPVQGMIRNPAASIPMRSSQPGQRTLOSOVANNIGSELEMANNG 1020  
 DB 961 VRYTCAATTSAAMNRPVQGMIRNPAASIPMRSSQPGQRTLOSOVANNIGSELEMANNG 1020  
 QY 1021 POYSOQOAPPNQTAPEPSSILPIDQASFASONROPPGSSPPDLCPHPAAESPSEGALL 1080  
 DB 1021 POYSOQOAPPNQTAPEPSSILPIDQASFASONROPPGSSPPDLCPHPAAESPSEGALL 1080  
 QY 1081 DQLYLALNFDGLIEIDALGIPELVSOQAADVDPQFSSQSDSNIMLEQKAPFPQOYASQ 1140  
 DB 1081 DQLYLALNFDGLIEIDALGIPELVSOQAADVDPQFSSQSDSNIMLEQKAPFPQOYASQ 1140  
 QY 1141 AQMAQGSYSPMODPNFHTMGORPSTATLRMQPRPLRPTGLVQONPNQLRLQALQRLAQ 1200  
 DB 1141 AQMAQGSYSPMODPNFHTMGORPSTATLRMQPRPLRPTGLVQONPNQLRLQALQRLAQ 1200

QY 1201 QNRPLMNOISVSNVNLTRPGVPTQAPINAOMLAQRORREILNOLROROMHOQQOQVQ 1260  
 DB 1201 QNRPLMNOISVSNVNLTRPGVPTQAPINAOMLAQRORREILNOLROROMHOQQOQVQ 1258  
 QY 1261 RTLMMGQGLNMTPSVAPSGMATSNPRILQANQOQFFPPRYGTSQQPDDGFTGAT 1320  
 DB 1259 RTLMMGQGLNMTPSVAPSGMATSNPRILQANQOQFFPPRYGTSQQPDDGFTGAT 1318  
 QY 1321 POSPLMSPRAHQSPMAQSOANPAYOAPSDINGNAGNMGNSMFSQSQSPHFGQAN 1380  
 DB 1319 POSPLMSPRAHQSPMAQSOANPAYOAPSDINGNAGNMGNSMFSQSQSPHFGQAN 1378  
 QY 1381 TSMYSNNMNTSVSMATVTGMSMNMOMTGOISMTSVTSSTGLSSMGPQVNDPALRG 1440  
 DB 1379 TSMYSNNMNTSVSMATVTGMSMNMOMTGOISMTSVTSSTGLSSMGPQVNDPALRG 1438  
 QY 1441 NLFNPOLPGMDMKOEEDTRKXC 1464  
 DB 1439 NLFNPOLPGMDMKOEEDTRKXC 1462  
 RESULT 3  
 NC02\_RAT STANDARD; PRT: 1465 AA.  
 AC Q9W019;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Nuclear receptor coactivator 2 (Transcriptional intermediary factor 2).  
 GN NCOA2 OR TIF2.  
 OS Rattus norvegicus (Rat).  
 OC Chordata; Vertebrata; Euteleostomi; Mammalia; Metazoa; Chordata; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxId=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Liver;  
 RX MEDLINE=98414616; PubMed=9742117;  
 RA Leers J., Treuter E., Gustafson J.-A.;  
 RT "Mechanistic principles in NR box-dependent interaction between nuclear hormone receptors and the coactivator TIF2.";  
 RL Mol. Cell. Biol. 18:6001-6013(1998).  
 CC -1- FUNCTION: TRANSCRIPTIONAL COACTIVATOR FOR STEROID RECEPTORS AND NUCLEAR RECEPTORS. COACTIVATOR OF THE STEROID BINDING DOMAIN (AF-2) BUT NOT OF THE MODULATING N-TERMINAL DOMAIN (AF-1).  
 CC -1- SUBCELLULAR LOCATION: Nuclear.  
 CC -1- SIMILARITY: CONTAINS 1 PAS (PER-ARNT-SIM) DIMERIZATION DOMAIN.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AF136943; AAD24587.1; -  
 CC InterPro: IPR001092; HLH\_dim.  
 CC InterPro: IPR001610; PAC.  
 CC InterPro: IPR000014; PAS.  
 CC Pfam: PF00889; PAS.1.  
 CC SMART: SM00353; HLH.1.  
 CC SMART: SM00086; PAC.1.  
 CC SMART: SM00091; PAS.1.  
 CC PROSITE: PSS0112; PAS.1.  
 CC Transcription regulation; Activator; Nuclear protein.  
 FT DOMAIN 119 183 PAS.  
 FT DOMAIN 1254 1260 POLY-QLN.  
 SQ SEQUENCE 1465 AA; 159435 MW; 36625B573EB0B39C CRC64;

Query Match

93.8%; Score 7160.5; DB 1; Length 1465;

Best Local Similarity 92.8%; Pred. No. 0;  
 Matches 1359; Conservative 54; Mismatches 51; Indels 1; Gaps 1;  
 QY 1 MSGGENTSPSRAEMRKRECPDQIGSPKRMTERKREKKEKYEELAEELIFANNDI 60  
 DB 1 MSGGENTSPSRAEMRKRECPDQIGSPKRMTERKREKKEKYEELAEELIFANNDI 60  
 QY 61 DNFNFKPKCAILKEVYKQIQRIKEDEKAAANIDEVQKSDVSTGGVLDKALGPM 120  
 DB 61 DNFNFKPKCAILKEVYKQIQRIKEDEKAAANIDEVQKSDVSTGGVLDKALGPM 120  
 QY 121 EALDGEFFVNLGNVYFVSENVYQYIRYQEBELANKSVYSLHVGDFEYVKNLPKST 180  
 DB 121 EALDGEFFVNLGNVYFVSENVYQYIRYQEBELANKSVYSLHVGDFEYVKNLPKSM 180  
 QY 181 VNGSMGEPBRNSHFNCRMLVYKPLDSEEGHNDQEAHQYETMQCAVSAOPSIKE 240  
 DB 181 VNGSMGEPBRNSHFNCRMLVYKPLDSEEGHNDQEAHQYETMQCAVSAOPSIKE 240  
 QY 241 EGEDLOSLICVARYPMKEPVLPSSESFTTRDLOGKITSLDSTYMAAMKPGMEDLV 300  
 DB 241 EGEDLOSLICVARYPMKEPVLPSSESFTTRDLOGKITSLDSTYMAAMKPGMEDLV 300  
 QY 301 RCIQKFAHGEBSVYAKRHHEVYRGLAFSQTIRSLSGTLVYAOTKSLIRSQ 360  
 DB 301 RCIQKFAHGEBSVYAKRHHEVYRGLAFSQTIRSLSGTLVYAOTKSLIRSQ 360  
 QY 361 TNBPOLVLSHMLHREBNVCMNPDLTGOTMGPLNLTSSNPAHOLCGNGGODPTLS 420  
 DB 361 TNBPOLVLSHMLHREBNVCMNPDLTGOTMGPLNLTSSNPAHOLCGNGGODPTLS 420  
 QY 421 SNINFPINGPKMGMPGMRGSGGNHVSQMATTPOGSNYALKMNSPQSGPMGP 480  
 DB 421 SNINFPINGPKMGMPGMRGSGGNHVSQMATTPOGSNYALKMNSPQSGPMGP 480  
 QY 481 QPTSMKSPRHRMSPVAGSRIRPSPQSPAGSLHSPGVCSSTGNSTYNSLNLQAL 540  
 DB 481 QPTSMKSPRHRMSPVAGSRIRPSPQSPAGSLHSPGVCSSTGNSTYNSLNLQAL 540  
 QY 541 SEGHVSLGSSLASPDLMKMNLTQNSPVNMNPPPLSKKSGTSDSCDCELYEPSEGTGGA 600  
 DB 541 SEGHVSLGSSLASPDLMKMNLTQNSPVNMNPPPLSKKSGTSDSCDCELYEPSEGTGGA 600  
 QY 601 ESSCHPEQKERTNDPMLPPAVSSERADQSRHDSKQTKLQLLTTKSDOMEPSPLASS 660  
 DB 601 ESSCHPEQKERTNDPMLPPAVSSERADQSRHDSKQTKLQLLTTKSDOMEPSPLASS 660  
 QY 661 LSPDNKDSSTGLSGSGSTHGTSLKEKHILHRLLODSSSPVDLAKLAETGKELNDSS 720  
 DB 661 LSPDNKDSSTGLSGSGSTHGTSLKEKHILHRLLODSSSPVDLAKLAETGKELNDSS 720  
 QY 721 STAPGSEVTIKQEPVSPKKKENALRYLDDKDTKDGLPEITPKLERLDSKTPASNTK 780  
 DB 721 STAPGSEVTIKQEPVSPKKKENALRYLDDKDTKDGLPEITPKLERLDSKTPASNTK 780  
 QY 781 LIAMKTEKEEMSEFEPDQPSGLDNEELIDDLQNSQLPOLFPDTRGAPAVDQKAI 840  
 DB 781 LIAMKTEKEEMSEFEPDQPSGLDNEELIDDLQNSQLPOLFPDTRGAPAVDQKAI 840  
 QY 841 NDLMQLTAENSPYTPVGAQCTALRIGSTFNNRPPQGLGLLNNUNPLDITIQSPGAG 900  
 DB 841 NDLMQLTAENSPYTPVGAQCTALRIGSTFNNRPPQGLGLLNNUNPLDITIQSPGAG 900  
 QY 901 PPPPIRNSPYSVIPOPGMGNGOIMGNLGNSSFTGMGNSASRPMTSGEAPQSSA 960  
 DB 901 PPPPIRNSPYSVIPOPGMGNGOIMGNLGNSSFTGMGNSASRPMTSGEAPQSSA 960  
 QY 961 VRYTCAATTSAMNRPVQGGIRNPPASIPMRPSSQPPQRLQSOVMNIGPSELENNMG 1020  
 DB 961 VRYTCAATTSAMNRPVQGGIRNPPASIPMRPSSQPPQRLQSOVMNIGPSELENNMG 1020  
 QY 1021 PYSQQAAPPNQTAPWPESTLPTDQASFASQNRPGSSPDDLCPHPAESPDEGALL 1080  
 DB 1021 PYSQQAAPPNQTAPWPESTLPTDQASFASQNRPGSSPDDLCPHPAESPDEGALL 1080

```
Db 1021 POLYNOQAAPPNQTAFWPESIPIIDQAFSGSONRHPFGSSPDDLCPHPAASPSDEGALL 1080
Oy 1081 DOLYALALNFDLEIEDIALGIPELVSOQAVDPFOFSODSNIMLEOKAPVEPOQYASQ 1140
Db 1081 DOLYALALNFDLEIEDIALGIPELVSOQAVDPFOFSODSNIMLEOKAPVEPOQYASQ 1140
Oy 1141 AQMAAGSYSPMODPNEHTMGORPSYATILMOBPRLPTGLVQNPOLRLQJHRLQAO 1200
Db 1141 AQMAAGSYSPMODPNEHTMGORPSYATILMOBPRLPTGLVQNPOLRLQJHRLQAO 1200
Oy 1201 QNRQPLMAQISVSNVNTLRGCVPTQAPINQAQRLNLAQORRETLNHLRQRMHQOQOYQ 1260
Db 1201 QNRQPLMAQISVSNVNTLRGCVPTQAPINQAQRLNLAQORRETLNHLRQRMHQOQOYQ 1260
Oy 1261 RFLMRRGGGLMNTPSMAVAPSGPATMSNPRIQANAOQPPRPNTGISQOPRGFTGAT 1320
Db 1261 RFLMRRGGGLMNTPSMAVAPSGPATMSNPRIQANAOQPPRPNTGISQOPRGFTGAT 1320
Oy 1321 POSPLMSPRMAHTOSPRMAQSOANPAYQAPSDINGMAQNMGNSMFSSQSPPHFGQOAN 1380
Db 1321 POSPLMSPRMAHTOSPRMAQSOANPAYQAPSDINGMAQNMGNSMFSSQSPPHFGQOAN 1380
Oy 1381 TSMYSNNNNINVSATNTGMSNNQMTGQISMTSVTSYSTGLSMGPEQVNDPALRG 1440
Db 1381 TSMYSNNNNINVSATNTGMSNNQMTGQISMTSVTSYSTGLSMGPEQVNDPALRG 1440
Oy 1441 NLF-PNOLPGMDMKOEGDTRKYC 1464
Db 1441 SLETTNOLPGMDMKOEGDTRKYC 1465

RESULT 4
NC02_XENLA
ID NC02_XENLA STANDARD: PRT: 1516 AA.
AC Q9W705;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Nuclear receptor coactivator 2 (Transcriptional intermediary factor 2) (XNIF2).
GN NC02 OR TIF2.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
OC Xenopodidae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=20171035; PubMed=10704837;
RA de la Calle-Mustienes E., Gomez-Skarmeta J.L.;
RT "XNIF2, a Xenopus homologue of the human transcription intermediary
RT factor, is required for a nuclear receptor pathway that also
RT interacts with CBP to suppress Brachyury and Xmyob.";
RL Mech. Dev. 91:119-129 (2000).
CC -1- FUNCTION: TRANSCRIPTIONAL COACTIVATOR FOR STEROID RECEPTORS AND
CC NUCLEAR RECEPTORS. COACTIVATOR OF THE STEROID BINDING DOMAIN
CC (AF-2) BUT NOT OF THE MODULATING N-TERMINAL DOMAIN (AF-1).
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED HOMOGENEUSLY DURING LATE BLASTULA-
CC NOTOCHORD.
CC -1- SIMILARITY: CONTAINS 1 PAS (PER-ARNT-SIM) DIMERIZATION DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AJ243119; CAB45389.1;
CC InterPro: IPR001092; HLH_dlm.
```

```
DR InterPro: IPR000014; PAS.
DR Pfam: PF00989; PAS; 1.
DR SMART: SM00353; HLH; 1.
DR SMART: SM00091; PAS; 1.
DR PROSITE: PS00112; PAS; 1.
KW Transcription regulation; Activator; Nuclear protein.
FT DOMAIN 116 180 PAS.
FT DOMAIN 1237 1273 POLY-GLN.
SQ SEQUENCE 1516 AA; 166156 MW; 09851C00A8439A4A CRC64;

Query Match 69.1%; Score 5276.5; DB 1; Length 1516;
Best Local Similarity 69.0%; Pred. No. 8-7e-233;
Matches 1058; Conservative 124; Mismatches 225; Indels 127; Gaps 19;

Oy 4 MGEINTSDPSRAETRRKREKEDDOLGSPKRNTEKRRNEQENKYTEIELAEILFANFNIDNF 63
Db 1 MGEINTSDPSRAETRRKREKEDDOLGSPKRNTEKRRNEQENKYTEIELAEILFANFNIDNF 60
Oy 64 NFKPKCATLKETVKQIRQIKOEKAAANIDEVOKSDVSTGGGYIDKDALGPMYLEAL 123
Db 61 NFKPKCATLKETVKQIRQIKOEKTAANDEVOADYSTGGYIDKDALGPMYLEAL 120
Oy 124 DGEFFVNLGNVNVSENVNTQYLRYNOBELANKSVYSILAHGDTHEFYKNLIPKSYNG 183
Db 121 DGEFFVNLGNVNVSENVNTQYLRYNOBELANKSVYSILAHGDTHEFYKNLIPKSYNG 180
Oy 184 GSWSGEPPRRNSHTNRCMLVAPLPDSEEGHDNOBAHOKYTEMOCFAVQSPKSIKEGE 243
Db 181 -----VPRRNSHTNRCMLVAPLMMECEERHDGQETHOKYESMOCFAVQSPKSIKEGE 234
Oy 244 DLQSLICVARRVPKERVLPSPSEFTTRDOLQKITSLDTSTRAAMKPGMEDLVRC 303
Db 235 DFQSLICVARRVPKERVLPSPSEFTTRDOLQKITSLDTSTRAAMKPGMEDLVRC 294
Oy 304 IOKFHAOHEGESVYAKRHHNEVLROGLAFSOLYRFSSLDGTLVAAQTOSKILRSQTTNE 363
Db 295 IOKFHAOHEGESVYAKRHHNEVLROGLAFSOLYRFSSLDGTLVAAQTOSKILRSQTTNE 353
Oy 364 POLVLSLMLHREQNVCMNPDLTGOTMGKPLNPISNSPAHQALCSGPNQDMLSSNI 423
Db 354 PLVLSLMLHREQNVCMNPDLTGOTMGKPLNPISNSPAHQALCSGPNQDMLSSNI 413
Oy 424 NPPINGPKROMPMRGREGSGGMNHVSGMQATTPGGSYVALAKMNSPOSSGPMNPGORT 483
Db 414 NPPINGPKROMPMRGREGSGGMNHVSGMQATTPGGSYVALAKMNSPOSSGPMNPGORT 473
Oy 484 SMLSPRRHMSPGVACSPRIPTQSPAGSLHPVGVCSSTGNSHTNSLNAALALSG 543
Db 474 SMLSPRRHMSPGVACSPRIPTQSPAGSLHPVGVCSSTGNSHTNSLNAALALSG 533
Oy 544 HGVSLGSSIASPDLKMLGNLQNSPVNMNPPRLSKMGLSKDGFGLYGEPSSECTQAESS 603
Db 534 QG-PLAPPLSSPDLKMLGNLQNSPVNMNPPRLSKMGLSKDGFGLYGEPSSECTQAESS 592
Oy 604 CHRGOKETNDNLPRAVSEADQSRHDSKQGTLLQLLTSSDQMEPPLAS-SLS 662
Db 593 CHSNQKDCGE-NLSSVY--DKTEGQSRLLDGKGGQKLLKLTTSDDQMEPPLAS-SLS 649
Oy 663 DPNKSTGSLP-----SGSHGTSLKEKHILRLLODSSSPVDLKLTLAEAGKLSDES 719
Db 650 DPNKSTGSLP-----SGSHGTSLKEKHILRLLODSSSPVDLKLTLAEAGKLSDES 709
Oy 720 SSTAPGEVTTIQEVPSPKKENALLRYLDKDTKDGLEPITPKLERLSDKTPASNT 779
Db 710 NSTGSGSEVTTIQEVPSPKKENALLRYLDKDTTDD--NVADITPKLERLSDKTPASNT 768
Oy 780 KLIANKTRKEEVSFPGQPGSELNLELDLIDLQNSQLPOLFPRTTRGAPAGSDKAI 839
Db 769 KLSAFAKEEVEFPHGHTQPGSDPNLDELIDLQNSQLSOLFSDTR--HDGNSADKQAI 826
Oy 840 INDLQQLTAENSVPVPGAK--TALRISQST-FNNPRGQLGRLLPNONLPDITLQSP 897
Db 830 INDLQQLTAENSVPVPGAK--TALRISQST-FNNPRGQLGRLLPNONLPDITLQSP 817
```

Db 827 MNDLMOLAGENSTGLPAPMAOKORMLRMQNGNNGENSOALAOLGR-LPNOULPLDIFHQSQA 885  
 Qy 898 GAGPPPIRNSSPYSVYIPQGMNGNOGMIGNOGNLGNSSTGMIGNSASRPTMPSGEWAPQ 957  
 Db 886 SASPSFQMRSSGYTYTPQSGVITNNOAMKSSQGNVRSSGITYGVNPRRLPRKPGDMGSO 945  
 Qy 958 SSAVRVTCATTSAMNRPVQGMIRNPAASIPMRSSQPGQKOTLOSQVNNIGPSELENN 1017  
 Db 946 ASAVRPACPTTSTAMNR---HDMRSRPTASTIPMRPGSQVCPROYLQSAVNMSSSELDNM 1002  
 Qy 1018 MGSPQYSQOQAPPNQTPWPESTILPIDQAFSAONQRPRESSPDLLCPHRAESPDG 1077  
 Db 1003 ISGPQYQQQAPPNQTPWPNRILTLTEQPSFNQNRQPFSSPADDLICQPIVSESPADG 1062  
 Qy 1078 ALLDQYTLARNPDGLLEIRALGIPELVQSQAADPEQSSODSNIMLROKAPVFEQOY 1137  
 Db 1063 NLLDQYTLARNPDGLLEIRALGIPELVQSQAADPEQSSODSNIMLROKAPVFEQOY 1122  
 Qy 1138 ASQAQMAQSGYSFPMQDPNFTMGQRPSTYATLRMQPRPGLRPTGLVQNPQRLQLOHRL 1197  
 Db 1123 ANQGMAMQNSYQPMQDPNFTMGQRPSTYATLRMQPRPGLRPTGLVQNPQRLQLOHRL 1182  
 Qy 1198 QAQONQPLMNOISNYSNVNLTLPQGP---TQAPINAQMLAQORELLINQHLROKQMH 1253  
 Db 1183 QA-QNRQQLMNPINNVSNNMLAMRPQVQGLRQGPINQMLAQORELLINQHLROKQMH 1241  
 Qy 1254 QQQQVQO-----RTLMRQGLINMTPSYAPSGMPATMS 1287  
 Db 1242 QQQQVQO-----RTLMRQGLINMTPSYAPSGMPATMS 1301  
 Qy 1288 NRPICQANAOQFPPPPY-----RTLMRQGLINMTPSYAPSGMPATMS 1305  
 Db 1302 SPRTQSTQOQFPPPPYGTGIPSPFPPTSPFSPVPPSPQSLSHSLHSGQMLNAG 1361  
 Qy 1306 -----GISQDPDPGFTGATTQSPPLMSRMAHOSPM 1338  
 Db 1362 IMGSMGQYPPVNPQMHNAFOFANSQMSQSDPGCTGATTQSPPLMSRMAHOSPM 1421  
 Qy 1339 QOQSOANPAYQAPSDINGMAQNGNMGSMFQSPHFGQOQANTSMY-SNNMNINVSMTN 1397  
 Db 1422 QOQSOANPAYQ-SELNMGAGNPNAGNSMFQSPHFGQOQANTSMY-SNNMNINVSMTN 1479  
 Qy 1398 TCGSMSSNOMTGOISMTSVTSVTSGLSSMGPEQ 1431  
 Db 1480 GNGMNNMNOMTGOISMTSVTSVTSGLSSMGPEQ 1513  
 RESULT 5  
 SIMA\_DROME STANDARD; PRT; 1507 AA.  
 ID 024167; Q9VNA5;  
 AC 15-DEC-1998 (Rel. 37, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Similar protein.  
 GN SIMA OR CG7951.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachyera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OC NCBL\_TaxID=7227;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96269413; Pubmed=8682312;  
 RA Nambu J.R., Chen W., Hu S., Crews S.T.;  
 RT "The Drosophila melanogaster similar bHLH-PAS gene encodes a protein  
 RT related to human hypoxia-inducible factor 1 alpha and Drosophila  
 RT single-minded.";  
 RL Gene 172:249-254 (1996).  
 RN (2)  
 RP SROUENCE FROM N.A.  
 RC STRAIN=BERKELEY;  
 RX MEDLINE=20196006; Pubmed=10731132;

RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers J.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abri J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Fester C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mallet B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Mostrelti A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Murthy D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacle J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner E., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster.";  
 RL Science 287:2185-2195(2000).  
 CC -1- FUNCTION: POSSIBLE DNA-BINDING TRANSCRIPTIONAL ACTIVATOR.  
 CC -1- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER  
 CC BHLH PROTEIN.  
 CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).  
 CC -1- TISSUE SPECIFICITY: UBICITOUSLY EXPRESSED IN THE EMBRYO.  
 CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF  
 CC TRANSCRIPTION FACTORS. HIGHEST TO HIF-1 ALPHA.  
 CC -1- SIMILARITY: CONTAINS 2 PAS (PER-ARNT-SIM) DIMERIZATION DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 1 PAS-ASSOCIATED C-TERMINAL (PAC) DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: U43090; AAC47303.1; -  
 DR EMBL: AE003772; AAF57008.2; -  
 DR Flybase: FBgn0015542; sima.  
 DR InterPro: IPR003015; HLH\_Myc.  
 DR InterPro: IPR001092; HLH\_dim.  
 DR InterPro: IPR001610; PAS.  
 DR InterPro: IPR000014; PAC.  
 DR Pfam: PF00989; PAS; 2.  
 DR Pfam: PF00989; PAS; 2.  
 DR SMART: SM00353; HLH; 1.  
 DR SMART: SM00086; PAC; 1.  
 DR SMART: SM00091; PAS; 2.  
 DR PROSITE: PS00038; HELIX\_LOOP\_HELIX; 1.  
 DR PROSITE: PS00112; PAS; 2.  
 KW Repeat: DNA-binding; Nuclear protein; Transcription regulation;  
 KW Activator; Coiled coil.  
 FT DNA\_BIND 72 85 BASIC DOMAIN.







OY 240 EEGEDLQSLCLVCARVRPMKERNVLPSSSEFTTRDLOGKITSLDTSRMAAMPKGMEDL 299  
 Db 265 E-----MCTV-----NDEFTSRSLSEWKLFLD---HRAPILIGYLPF 302  
 OY 300 VARCIOKHAQHEGSESVYAKRHHNEVLROGLAFSQITRFSLSDSTLYAAQTKSLINSQ 359  
 Db 303 EVLGTSGDYHYVADLEMLAKCHEH-LMAYGKSKCYRFLKGGQOMWLQTHYVITYHQ 361  
 OY 360 TTNEPOLVLSLHMLREQNVCMNPDGLGQWKGKPLNPSSNSPAHOLCSPNPQDMTL 419  
 Db 362 WNSRREFVTCHTVTVSYAEVRAERRELQ-----IEESLPETAADKSDSDSNDRI 412  
 OY 420 SSNINFPINGREOMKPMGRFGSGGNHNVSGMOATTPOGSNVALKMN--PSQSPGMN 478  
 Db 413 NT-----VSLKEALERDHS-----PTPSASSRSRSRSHAVDPST 451  
 OY 479 PCQ-PTSMLSPRHNSPVPAGSPRIPPSOFSPAGSLHSPVGCSTGNSHTNSLNL 537  
 Db 452 PTKIPTDNT-----PPROHLPA--HEKM-----VORRSFSSQINS- 487  
 OY 538 QALSEGHVGLSSLASPDLMKGNLQNSPVNMNPPPLS-----KMSLDS-KDCEGLYG 590  
 Db 488 -----QSVGSSLTGP--VMSQATNPPI--PGMSQFQPSAQDGMQHLKD----- 528  
 OY 591 EPSEGTQAESSCHPGQOKETNDPNLPAYSSERADG-QSRMLDSK-----GQTKLLQL 644  
 Db 529 -OLEQRTMTEANIH-RQOEELR--KIQEOQLMVHGQGLQWLPQSNGLNCGVQLSSG 584  
 OY 645 LTTSDQWEPPLASSLSDTKNDKSTGSLPGS-GSTHGTSLEKHKILHRLQDSSPPDL 703  
 Db 585 NSSNIQQLAPINMOQVVPTRNIOISGMNTGHGTT-----QHMIOQOTLOSTSTOSQ 637  
 OY 704 AKLTAETAKGKLSQSSSS--TAPGEVTIKQEPVSPKKENALLYLKDKDTKIGLPE 761  
 Db 638 NVLSHSGQOTSLPSCQSTQTLAPLNTVVISOPAGSM-----VQIPS 680  
 OY 762 TTPKLERLDSKTPASNTKLIAMKTEKEMSEPPDQ 798  
 Db 681 SMPQ-----NSTQSAAVTTF-----TQDRQIRFSGCQ 708

RESULT 7  
 ARNT\_DROME STANDARD; PRT; 644 AA.  
 ID ARNT\_DROME  
 AC 015945: 016167: 044082: 09VH12: Q24461;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Aryl hydrocarbon receptor nuclear translocator homolog (dARNT)  
 DE (tango protein) (Hypoxia-inducible factor 1 beta).  
 DE TGO OR ARNT OR HIF-1-BETA OR CG11987.  
 GN Drosophila melanogaster (Fruit fly).  
 OS Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 CC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN 11  
 RP SEQUENCE FROM N.A., FUNCTION, AND SUBUNIT.  
 RC TISSUE-Embryo.  
 RX MEDLINE=98040551; PubMed=9374395;  
 RA Onshiro T., Saigo K.;  
 RT "Transcriptional regulation of breathless FGF receptor gene by  
 RT binding of TRACHEALLESS/dARNT heterodimers to three central midline  
 RT elements in Drosophila developing trachea.";  
 RL Development 124:3975-3986(1997).  
 RN 12  
 RP SEQUENCE FROM N.A., FUNCTION, SUBUNIT, AND TISSUE SPECIFICITY.  
 RC TISSUE-Embryo;  
 RX MEDLINE=97427859; PubMed=9284047;  
 RA Zelzer E., Wapner P., Shilo B.-Z.;  
 RT "The PAS domain confers target gene specificity of Drosophila  
 RT bHLH/PAS proteins";  
 RL Genes Dev. 11:2079-2089(1997).

RN 13  
 RP SEQUENCE FROM N.A., FUNCTION, SUBUNIT, TISSUE SPECIFICITY, AND  
 RP DEVELOPMENTAL STAGE.  
 RC TISSUE-Embryo;  
 RX MEDLINE=98072332; PubMed=9409674;  
 RA Sonnenfeld M., Ward M., Nyström G., Mosher J., Stahl S., Crews S.;  
 RT "The Drosophila tango gene encodes a bHLH-PAS protein that is  
 RT orthologous to mammalian Arnt and controls CNS midline and tracheal  
 RT development";  
 RL Development 124:4571-4582(1997).  
 RN 14  
 RP SEQUENCE FROM N.A., FUNCTION, AND DEVELOPMENTAL STAGE.  
 RC STRAIN-CANTON-S; TISSUE=Embryo, Pupae, and Head;  
 RX MEDLINE=20050848; PubMed=10581393;  
 RA Ma E., Haddad G.G.;  
 RT "Isolation and characterization of the hypoxia-inducible factor 1beta  
 RT in Drosophila melanogaster";  
 RL Brain Res. Mol. Brain Res. 73:11-16(1999).  
 RN 15  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BERKELEY;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Branton R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abell J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,  
 RA Borokova D., Botchan M.R., Bouck J., Brockstein P., Brottier P.,  
 RA Butts K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Folsler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Honick J.,  
 RA Hostin D., Houston K.A., Howland T.D., Wei M.-H., Ibegwam C.,  
 RA Jaisl M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Paclob J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhu D., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster";  
 RL Science 287:2185-2195(2000).  
 RN 16  
 RP SEQUENCE OF 586-623 FROM N.A.  
 RX MEDLINE=87051745; PubMed=2877746;  
 RA Frigerio G., Burri M., Bopp D., Baumgartner S., Noll M.;  
 RT "Structure of the segmentation gene paired and the Drosophila PRD gene  
 RT set as part of a gene network";  
 RL Cell 47:735-746(1986).  
 CC -1- FUNCTION: TGO/TRH HETERODIMERS ARE INVOLVED IN THE CONTROL OF  
 CC BREATHLESS EXPRESSION. PLAYS A ROLE IN THE CELLULAR OR TISSUE  
 CC RESPONSE TO OXYGEN DEPRIVATION.  
 CC -1- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER  
 CC BHLH PROTEIN. HETERODIMER WITH AHR, TRH OR SIM.  
 CC -1- TISSUE SPECIFICITY: AT STAGE 11, EXPRESSION IS DETECTED IN



RL Cell 89:641-653(1997).  
 RN [3]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN-129/SV;  
 RA Wilsbacher L.D., Sangoram A.M., Antoch M.P., Takahashi J.S.;  
 RT "The mouse Clock locus: Sequence and analysis of 204 kb from mouse  
 chromosome 5.";  
 RL Submitted (May-2000) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RN INTERACTION WITH BMAL1.  
 RP MEDLINE-98279137; PubMed-9616112;  
 RX Gekakis N., Staknis D., Nguyen H.B., Davis F.C., Wilsbacher L.D.,  
 RA King D.P., Takahashi J.S., Weitz C.J.;  
 RT "Role of the clock protein in the mammalian circadian mechanism.";  
 RL Science 280:1564-1569(1998).  
 CC -1 FUNCTION: CIRCADIAN REGULATOR THAT ACTS AS A TRANSCRIPTION FACTOR.  
 CC CLOCK-BMAL1 HETERODIMERS BIND TO AN E-BOX ELEMENT (3'-CACGCG-5')  
 CC THEREBY ACTIVATING TRANSCRIPTION OF PER1, AND POSSIBLY OF OTHER  
 CC CIRCADIAN CLOCK PROTEINS. MUTANT CLOCK AND BMAL1 FORM HETERODIMER  
 CC THAT BIND DNA, BUT FAIL TO ACTIVATE TRANSCRIPTION. IN HOMODIMERS  
 CC CLOCK MUTANTS, THE CIRCADIAN PERIOD IS INCREASED FROM 3 TO 4 HOURS  
 CC AND USUALLY THE CIRCADIAN RHYTHMICITY IS LOST AT CONSTANT  
 CC DARKNESS. EXPRESSION OF CLOCK IS ALSO REDUCED.  
 CC -1 SUBUNIT: HETERODIMER WITH BMAL1 AND LESS EFFICIENTLY WITH ARNT  
 CC AND ARNT2. HETERODIMERS WITH ARNT OR ARNT2 BIND POORLY TO THE E-  
 CC BOX MOTIF.  
 CC -1 TISSUE SPECIFICITY: EXPRESSED EQUALLY IN BRAIN, EYE, TESTES,  
 CC OVARIES, LIVER, HEART, LUNG, KIDNEY. IN THE BRAIN, EXPRESSION IS  
 CC ABUNDANT IN THE SUPRACHIASMATIC NUCLEI (SCN), IN THE PYRIFORM  
 CC CORTEX, AND IN THE HIPPOCAMPUS. LOW EXPRESSION THROUGHOUT THE REST  
 CC OF THE BRAIN. EXPRESSION DOES NOT APPEAR TO UNDERGO CIRCADIAN  
 CC OSCILLATIONS.  
 CC -1 DOMAIN: CONTAINS A GLN-RICH C-TERMINAL DOMAIN WHICH CLOCK  
 CC CORRESPOND TO THE TRANSACTIVATION DOMAIN. IN MUTANT CLOCK,  
 CC DELETION OF THIS REGION LEADS TO AN INCREASED CIRCADIAN PERIOD  
 CC FROM 3 TO 4 HOURS AS WELL AS TO THE LOSS OF CIRCADIAN RHYTHMICITY  
 CC AND ALTERED LIGHT RESPONSE.  
 CC -1 DISEASE: DEFECTS IN CLOCK AFFECT TWO PROPERTIES OF THE CIRCADIAN  
 CC SYSTEM: THE LENGTH OF THE FREE-RUNNING PERIOD AND THE PERSISTENCE  
 CC OF CIRCADIAN RHYTHMICITY IN CONSTANT DARKNESS.  
 CC -1 SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF  
 CC TRANSCRIPTION FACTORS.  
 CC -1 SIMILARITY: CONTAINS 2 PAS (PER-ARNT-STM) DIMERIZATION DOMAINS.  
 CC -1 SIMILARITY: CONTAINS 1 PAS-ASSOCIATED C-TERMINAL (PAC) DOMAIN.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation at  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AF000998; AAC53200.1; -  
 DR EMBL; AF146793; AAB30565.1; -  
 DR MGD; MGI:99698; Clock.  
 DR InterPro; IPR003015; HLH\_MYC.  
 DR InterPro; IPR001092; HLH\_dim.  
 DR InterPro; IPR001067; Nucleinslocator.  
 DR InterPro; IPR001610; PAC.  
 DR InterPro; IPR000014; PAS.  
 DR Pfam; PF00785; PAC; 1.  
 DR Pfam; PF00989; PAS; 2.  
 DR PRINTS; PR00785; NCTRNSLOCATR.  
 DR SMART; SM00353; HLH; 1.  
 DR SMART; SM00086; PAC; 1.  
 DR SMART; SM00091; PAS; 2.  
 DR SMART; SM00509; TFS2N; 1.  
 DR PROSITE; PS00038; HELIX\_LOOP\_HELIX; 1.  
 DR PROSITE; PS50112; PAS; 2.  
 DR Transcription regulation; Nuclear protein; Repeat; Biological rhythms;  
 DR DNA-binding; Alternative splicing.

FT	DNA_BIND	35	47	BASIC DOMAIN.
FT	DOMAIN	48	85	HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).
FT	DOMAIN	107	177	PAS 1.
FT	DOMAIN	262	332	PAS 2.
FT	DOMAIN	484	855	GLN-RICH.
FT	DOMAIN	740	745	POLY-GLN.
FT	DOMAIN	751	759	POLY-GLN.
FT	DOMAIN	762	769	POLY-GLN.
FT	DOMAIN	828	837	POLY-GLN.
FT	DOMAIN	514	564	IMPLICATED IN THE CIRCADIAN RHYTHMICITY.
FT	VARSPLIC	484	513	MISSING (IN SHORT ISOFORM).
FT	SEQUENCE	855 AA:	96393 MM:	9664D947049742F4 CRC64;
Query Match 4.1%; Score 309.5; DB 1; Length 855;				
Best Local Similarity 21.7%; Pred. No.3.4e-07;				
Matches 150; Conservative 107; Mismatches 254; Indels 181; Gaps 27;				
QY	31	KNTTEKRNDEENKTYIEELAEIIFANFNPDIDNENFKPKCALIKETVKQIQKOEKAA	90	
DB	39	RNKSEKKRRDQFNVLKELGSMLPENAR-----KMDKSTVLQKSIDFLR--KHKETTA	89	
QY	91	AANIDEVCKSDVSSGGQGVIDKDALGPMLEALDGEFFVNVLEGNVWVSEVWVQYLKYN	150	
DB	90	QGDASEI--RQMKKP---FLSNEPTQMLELIDDFPLAIMDGSIIIVSEVSLLEHL	145	
QY	151	QEELMKNSVYSILHGDHTEFEVKNLPKSIYVGSMSGEPPR--RNSHTFNCRL-----	203	
DB	146	PEDIVDQSTFNNIPGEHSEVYK--ILSHLILESDDLTEFYELKSKNQLEFCCLMLGCTIDP	204	
QY	204	-----VKPLPDEEEGHNDN-----QEAHQ--KYETMOCFA---VSPQKSIR	239	
DB	205	KEPSTEYEVYRFIGNKESLITSYSTHNGEPEGIQTHRPSYEDRVCFAVATVRLAPQFIK	264	
QY	240	EEGEDLOSLICVARNVPMKERVLPBSSSTFTRDDLOGKITSLDTSTMRAMKQGWEDL	299	
DB	265	E-----MGTV-----EEP-----NEEFTSRSLMKLFLELD--HRAPIIGLYPF	302	
QY	300	VRRCIQKFAHQEGESVSAYAKRHHEHEVLKQGLAFQIYRFSLSDDTLVAQTKSKLIRSQ	359	
DB	303	EVLGTSGYVYHVDDELNAKCHEN--LMQYKGRKCYRFLTKGQWIMLQTHYITYHQ	361	
QY	360	TINEPOLVISLHMLREQNVCMYNDLTGQTMKPLNPITSSNSPAHQALCSGNPGQDML	419	
DB	362	MNSREFIYCTHTVSYAEVRAERRELG-----IEESLPETAADKSDSGSDNRI	412	
QY	420	SSNINPFIQREQMGMPRGFRGSGGMNHYSGMATTPOGSNYVALKANS--PSQSPGMN	478	
DB	413	NT-----VSLKEALERFDHS-----PTPSASRSRSRSSHVAVSDPST	451	
QY	479	PGQ-PTSMLSP-----RHHRSPGVASPRIPRPSQFPAGSLSPVVCSSST	523	
DB	452	PKIKTIPDITSPRRQILPAHEKMTQKRSSFSSQTSINSQVSGBELTPRAHSQANLPI--PQ	509	
QY	524	GNSHSTYNSSLNALQALSE-----GHGVSILGSSL	552	
DB	510	GMSQFQFSQDLQAMQHLKDQLQEQTRMIEANIHQOEELRKIQEQDLQWVHGQGLQMFQO	569	
QY	553	ASPDLMKMLQNS-----PVNANRP--PLSKMSLSDSKDCFGLYGPESBQ-----	595	
DB	570	SNRGLNFGSVQLSSGNSNIQQLTPRNMGQOVVPAQVOS-----GHISTQGMHIQ	619	
QY	596	-----TTGAQESSCHPGEOKETMDPMLPRA	620	
DB	620	QQTLDSTSTQSQSQSVMSGHSQQTSLSPQTPS	651	
RESULT 9				
NPA2_MOUSE				
ID	NPA2_MOUSE	STANDARD:	PRT:	816 AA.
AC	P97460:			
DT	15-DEC-1998 (Rel. 37, Last sequence update)			
DT	15-DEC-1998 (Rel. 37, Last sequence update)			

```

DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Neuronal PAS domain protein 2 (Neuronal PAS2).
GN NPAS2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=97165088; PubMed=9012850;
RA Zhou Y.-D., Barnard M., Tian H., Li X., Ring H.Z., Francke U.,
RA Shelton J., Richardson J., Russell D.W., McKnight S.L.,
RT "Molecular characterization of two mammalian DHLH-PAS domain proteins
RT selectively expressed in the central nervous system.";
RL Proc. Natl. Acad. Sci. U.S.A. 94:713-718(1997).
CC -SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER
CC BHLH PROTEIN. INTERACTS WITH HSP90.
CC -SUBCELLULAR LOCATION: Nuclear (potential).
CC -TISSUE SPECIFICITY: IN BRAIN, EXCLUSIVELY NEURONAL. ALSO FOUND IN
CC SPINAL CORD, AND IN A LESSER EXTENT IN COLON, SMALL INTESTINE AND
CC UTERUS.
CC -1- DEVELOPMENTAL STAGE: FIRST DETECTED 3 DAYS AFTER BIRTH.
CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
CC TRANSCRIPTION FACTORS.
CC -1- SIMILARITY: CONAINS 2 PAS (PER-ARNT-SIM) DIMERIZATION DOMAINS.
CC -1- SIMILARITY: CONAINS 1 PAS-ASSOCIATED C-TERMINAL (PAC) DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announcement/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: U77969; AAB47249.1; .
DR MGD: MGI:109232; NPas2.
DR InterPro: IPR003015; HLH_MYC.
DR InterPro: IPR001092; HLH_dlm.
DR InterPro: IPR001610; PAC.
DR InterPro: IPR000014; PAS.
DR Pfam: PF00010; HLH; 1.
DR Pfam: PF00785; PAC; 1.
DR Pfam: PF00989; PAS; 2.
DR SMART: SM00353; HLH; 1.
DR SMART: SM00086; PAC; 1.
DR SMART: SM00091; PAS; 2.
DR PROSITE: PS00038; HELIX_LOOP_HELIX; 1.
DR PROSITE: PS0112; PAS; 1.
KW Repeat; DNA-binding; Nuclear protein; Transcription regulation.
FT DNA_BIND 10 22
FT DOMAIN 23 60 BASIC DOMAIN.
FT DOMAIN 82 152 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).
FT DOMAIN 237 307 PAS 1.
FT DOMAIN 311 354 PAS 2.
FT SEQUENCE 816 AA; 90915 MW; 7E5CF0641CFDC1DD CRC64;
SQ
Query Match 3.8%; Score 289; DB 1; Length 816;
Best Local Similarity 18.3%; Pred. No. 2.7e-06;
Matches 211; Conservative 151; Mismatches 360; Indels 432; Gaps 42;
OY 4 MGENSDPSRATRRKKECPDQLGSPKRNTERKRRREGENKYTEELAEILFANFDINDF 63
DB 1 MEDERDKRRKRSRNK-----SEKKRRDDQNVILKELSSLPENTR----- 41
OY 64 NEKPPKCALIKETVQIRQIRKQKAAANIDVOKSDVSSFGQGVIRKDALGPMYLEAL 123
DB 42 --KMKQTYLLEKVIQGLQ--KHNYSACQTEICDIDQDKMPS-----FLNSEETQYLMLEAL 93
OY 124 DGEFFVNLGEGNVEVSENVQYLYRYNOEELNKKSVSYILHGVGHTPEFVKNLIPKSYING 183

```

Dd	94	DGEIVYVTTGGSLIIYVSDSTITPPLLGHLPADVMDQNLNLPRLQEHSEVYKILISMILVTD	155
Qy	184	GSWSGEBPRRRSHTEPCRMIL---VKP-----LPDSEEGHDN---	217
Dd	154	SPSDEFLKSNDDLEFYCHLRGLSLNKEPFEYIEIKFVGNFRSYNNVNSPSCNGDNTLS	213
Qy	218	QEAHQKKEWQCF---AVSQPSKIEBEBEDJQSLICARVAPKKEPVLSPSSSFTR	273
Dd	214	RPCHVPLKDVCFATVATRLATPQFLKE-----MCAV-----DEPL---EEFYSR	254
Qy	274	QDLOGKITSLDSTSTMRAMPGW---EDLVRCIOGFHQAQHEGESVYAKRHNEVTLPG	330
Dd	225	HSLEMKLFLFD---HRAPIITGLPREVLGTSQYNYTH---DDELLARHCHQIMFG	307
Qy	331	LAFSQAIFYSLSDGTLYAAQTKSLKIRSQTNPEPOLVISLH-----MLHREONCYM	382
Dd	308	KGKSCCYRFLTKGGQOMIWLQTHYIITYHQMNSKPEFIVCTHSVSVADYVERROELALE	367
Qy	383	NPDLTGOTMGPRLNPJISSNPAHQALCSGNPDQMTLSSNIPFINGREQMGMMGRFG	442
Dd	368	DP-----PTEAH-----	374
Qy	443	GGSGMNVHSGMOATTPQGSNYVALKMNPSOSSPGMNPQPTSMLEPRHMSGVAGSPRI	502
Dd	375	-----HPSAVK-----EKDSSLEPPPPFAL-----DMGASGLPSS	405
Qy	503	PPPSFSPAGSLHSPVGVCSGTGNSHSTVNSLMALOALSEGHVSLGSSLASPLDKMGL	562
Dd	406	P-----SPSASRS-----SHKSSHTAMS-----	424
Qy	553	QNSVYVNMNPPJLSMGLSDSKDQCFGLYGP-SEGTGQAESESCHEQEKETNDPMLPVA	621
Dd	425	-----BETSPYTKLMEENSTTALPRATLPQELPQOG	456
Qy	622	SSEBADOQSRLHDSKG---QTKLLOLTLTKSDOMEPSPLAS-----SLSDNNKTSGL	672
Dd	457	LSQALMPTALHSSACDLTKOLLDSLPQTGLQSPAPVYTOPSAQSFMTQIKD-----	511
Qy	673	PGSGSTHGTSLKEHKHILHRLDQSSPYVDLAKTLATGAKDLSOESSSTAGSEFVTIQ	732
Dd	512	-----QEQYTRILQ-----ANIRMQ	528
Qy	733	EPVSPKKENNALRYLLDKDQTDYDGLPELTTPKLEKLSKTPASNTKLIAMKTEKEMS	792
Dd	529	ELHAKIOEOLQVQ-----DSVQWF-LOQPAVSLS	558
Qy	793	FEPPDQGSSELNDNEILLDDJQNSQLPQLEPDTPRGAPAGSVQDKALINDMLQLAENSP	852
Dd	559	FSSSTQRPAAQ-----QQLQQ-----RPAAPS---QPOLVYN-----	586
Qy	853	VTPGAOKTALRISOS-----TFNNRPPQGLRLLPNONLPL-DTLTQSPGAPGF	902
Dd	587	-TTPLOGQITQVINGHLARESNVISAGQKPKRRSQLLPAGGRSLSSLPQSFSTASVL	645
Qy	903	PPIRNSSPYVILQPGMMGNQMGIMQGNLGN-STGMTGNSASRPTMFGEMADQSSAV	961
Dd	646	PPGLSTLIATPQ-----DDSOCSPDEGHDROLRLILSQIPQPMMGSCDARQSEV	700
Qy	962	RVT-----CAATSAMRPVYGGGIRTPAASIPRRPSSQGOGQOTLOS	1004
Dd	701	SRTGRQVKAQASQVMPSPSHPTNSSASTPV---LLMGQAVLHDSFVPSRSPLQ----	753
Qy	1005	QVNNIGSELEMMNMGCPYOSQOQAPRNOPLR---MPESILPIDAQSFASQONRPGS	1058
Dd	754	-----PAQAQOQPPRLQAPRTSLHSEQFDSL-----LSTFSQQPGTIGY	793
Qy	1059	SPDDLCLCPHAAES	1072
Dd	794	AATQSTPPQPPRPS	807
RESULT	10		
NFT5_HUMAN			

ID NPT5\_HUMAN STANDARD; PRT; 1531 AA.  
AC 094916; 095693; 09UN18;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Nuclear factor of activated T cells 5 (T cell transcription factor  
DE NFAT5) (NF-AT5) (Tonicity-responsive enhancer-binding protein) (Tone-  
DE binding protein) (ToneBP).  
GN NFAT5 OR TONEBP OR KIA0827.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_taxid:9606;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORM C).  
RC TISSUE-Brain;  
RX MEDLINE-99156230; PubMed-10048485;  
RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Hirosewa M.,  
RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;  
RT "Prediction of the coding sequences of unidentified human genes. XII.  
RT The complete sequences of 100 new cDNA clones from brain which code  
RT for large proteins in vitro.";  
RL DNA Res. 5:355-364(1998).  
RN [2]  
RP SEQUENCE FROM N.A. (ISOFORM A).  
RC TISSUE-Brain;  
RX MEDLINE-99307389; PubMed-10377394;  
RA Lopez-Rodriguez C., Aramburu J., Rakehan A.S., Rao A.;  
RT "NFAT5, a constitutively nuclear NFAT protein that does not cooperate  
RT with Fos and Jun.";  
RL Proc. Natl. Acad. Sci. U.S.A. 96:7214-7219(1999).  
RN [3]  
RP SEQUENCE FROM N.A. (ISOFORM A).  
RX MEDLINE-99162641; PubMed-10051678;  
RA Miyakawa H., Woo S.K., Dahl S.C., Handler J.S., Kwon H.M.;  
RT "Tonicity-responsive enhancer binding protein, a rel-like protein that  
RT stimulates transcription in response to hypertonicity.";  
RL Proc. Natl. Acad. Sci. U.S.A. 96:2538-2542(1999).  
RN [4]  
RP SEQUENCE FROM N.A. (ISOFORM B).  
RC TISSUE-Brain;  
RA Lopez-Rodriguez C., Aramburu J., Rakehan A.S., Copeland N.G.,  
RA Gilbert D.J., Thomas S., Distche C., Jenkins N.A., Rao A.;  
RT "NFAT5: The NF-AT family of transcription factors expands in a new  
RT direction.";  
RL Cold Spring Harb. Symp. Quant. Biol. 64:517-526(1999).  
RN [5]  
RP SEQUENCE OF 675-1531 FROM N.A.  
RC TISSUE-Brain;  
RX MEDLINE-20029268; PubMed-10565538;  
RA Zuehlke C., Kiehl R., Johannsmeyer A., Grzeschik K.H., Schwainger E.;  
RT "Isolation and characterization of novel CAG repeat containing genes  
RT expressed in human brain.";  
RL DNA Seq. 10:1-6(1999).  
CC -1- FUNCTION: PLAYS A ROLE IN THE INDUCIBLE EXPRESSION OF GENES.  
CC -1- REGULATES HYPERTONICITY-INDUCED CELLULAR ACCUMULATION OF  
CC OSMOLYTES.  
CC -1- SUBUNIT: DOES NOT BIND WITH FOS AND JUN TRANSCRIPTION FACTORS. BUT  
CC MIGHT BE CAPABLE OF FORMING STABLE DIMERS WITH DNA ELEMENTS.  
CC -1- SUBCELLULAR LOCATION: Nuclear.  
CC -1- ALTERNATIVE PRODUCTS: 3 ISOFORMS; A, B AND C (SHOWN HERE); MAY BE  
CC PRODUCED BY ALTERNATIVE SPLICING.  
CC -1- TISSUE SPECIFICITY: HIGHEST LEVELS IN SKELETAL MUSCLE, BRAIN,  
CC HEART AND PERIPHERAL BLOOD LEUKOCYTES. ALSO EXPRESSED IN PLACENTA,  
CC LUNG, LIVER, KIDNEY, PANCREAS, SPLEEN, THYMUS, PROSTATE, TESTIS,  
CC OVARY, SMALL INTESTINE AND COLON.  
CC -1- SIMILARITY: BELONGS TO THE REL/DORSAL FAMILY.  
CC -1- CAUTION: REF.5 SEQUENCE DIFFERS FROM THAT SHOWN FROM POSITION 1165  
CC ONWARD DUE TO A FRAMESHIFT.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
DR EMBL; AB020634; BAA74850.1; -  
DR EMBL; AF089824; AAD18136.1; -  
DR EMBL; AF134870; AAD38360.1; -  
DR EMBL; 297016; CA009693.1; ALT\_FRAME.  
DR EMBL; AF163836; AAD48441.1; -  
DR MIM; 604708; -  
DR InterPro; IPR002909; IPT\_TIG.  
DR InterPro; IPR000451; REL.  
DR Pfam; PF01833; TIG; 1.  
DR SMART; SM00429; IPT; 1.  
DR PROSITE; PS01204; REL\_1; FALSE\_NEG.  
DR PROSITE; PS0254; REL\_2; 1.  
KW Transcription regulation; Activator; Nuclear protein; DNA-binding;  
KW Alternative splicing.  
FT DOMAIN 69 100 SER-RICH. DNA-BINDING.  
FT DOMAIN 293 300 POLY-GLN.  
FT DOMAIN 739 743 POLY-GLN.  
FT DOMAIN 879 888 POLY-GLN.  
FT DOMAIN 966 971 POLY-THR.  
FT DOMAIN 1248 1266 POLY-GLN.  
FT VARSPPLIC 1 76 MISSING (IN ISOFORM A).  
FT VARSPPLIC 1 47 MISSING (IN ISOFORM B).  
FT CONFLICT 1369 1369 E -> D (IN REF. 5).  
SQ SEQUENCE 1531 AA; 165764 MW; A68C6808BDBAF69E CRC64;  
  
Query Match 3.8%; Score 289; DB 1; Length 1531;  
Best Local Similarity 20.9%; Pred. No. 6.1e-06;  
Matches 292; Conservative 164; Mismatches 545; Indels 396; Gaps 68;  
OY 219 EAHQYETMOCFAVVSQPSIKKEGEDLSCLICVARVRPMKRPVLPSESEFTTQDLOG 278  
DB 319 ECHNEPVVLQVVGNGDSGRVKNPHG-FYQACRYTGNTTPCKE-----VDIEG 364  
OY 279 KITSLDTSTMRAMKPGME-DLVRRICQFNAHGESESVYAKRHNEV-LRQGLAFSQ- 335  
DB 365 -----TVIEVGIDRPSNNMGLAVDC-----VGILKLNADVEARIGIAGSK 406  
OY 336 -----IYRPSL--SDGTVAAGT-KSLIRSQTNRPOLY-LSLH--MLHRENNCYM 382  
DB 407 KSTRALVFRVNIIMRKDSFTLTQTPSSPILCTOPAGVPEILKSLHSCSVAGEEVR-- 464  
OY 383 NPDLTGQTMGRPLNPISSNSPAHQALCSGNPGQDWTLSNIFPINC---KEQGMIPMG 439  
DB 465 ---LIGKMFLOKTKYIPQENVSDEMSKSEALIDMELFHQNLIVKYVPRYHDQHTTLFVS 521  
OY 440 ---REGSGGGMNHVSGMOATTPGGSNYALKMNSPSSQSPGNNPOPTSMLSPRHRMS-- 493  
DB 522 VGIVYVTVNAGSHDQVPTTPPDPAAGALNVNKKIS---SPARPCSFEBAMKMKTT 578  
OY 494 -----PGVAGSPHLPSPQSPASLHSPVCVCSGNSHY--INSSLMALQALSE 542  
DB 579 GGNLDKVNIIIPALMTPLIPSSMIKSEDV--TPMEVTAEKRSSTIFKTKYSVGSQPTLE 636  
OY 543 -----GNG-VSISGLSLSPDLKMGNLONSPPVNNMPPLSKGSLDSKDCFLGYEPSEG 595  
DB 637 NISNINAGNGSPSSPSSHLPSSENEKQQLDQPKAYNPEL----- 675  
OY 596 TTGAQESSCHPG-----EOKETNDNLPVAVSSERADGQSR-LNDSKQOTKLLOLT 646  
DB 676 TTICQODISQCTFPVAVSASSQLPNSDALLQDATQFQTRERISREILQSDGVVVMISOLT 735  
OY 647 TKSDMEESPIL-----ASLSPTNND-STGSLPGSGSTHTGSLK 684  
DB 736 EASQDQDQSPLOEQAQTLQOOISSNIPSPNSVSQDQNTIOOLQAGSTGTAS----- 789  
OY 685 EKXKILHLLDSSSPVDLAKILTA-----TGKDLQESSSTPAGSEVTTIKO 732





FT CONFLICT 27 27 D -> G (IN REF. 1).  
FT CONFLICT 51 51 V -> M (IN REF. 1).  
FT CONFLICT 349 349 O -> S (IN REF. 2).  
FT CONFLICT 368 368 G -> GG (IN REF. 3).  
FT CONFLICT 374 374 P -> S (IN REF. 2).  
FT CONFLICT 393 393 P -> L (IN REF. 2).  
FT CONFLICT 489 502 MAMPPOVSGMCP -> NGQYATSGWVDVS (IN REF. 2).  
FT CONFLICT 634 634 E -> G (IN REF. 2).  
FT CONFLICT 924 924 P -> S (IN REF. 2).  
FT CONFLICT 1016 1016 O -> L (IN REF. 2).  
FT CONFLICT 1033 1033 O -> P (IN REF. 2).  
SQ SEQUENCE 1902 AA: 205946 MW: 9870A46F81062EAA CRC64:

Query Match 3.7%: Score 284.5; DB 1: Length 1902;  
Best Local Similarity 21.4%: Pred. No. 1.3e-05;  
Matches 289; Conservative 126; Mismatches 450; Indels 483; Gaps 74;

OY 205 KPLPD--SEEGHNDQEAHOKET---MQCF-----AVGQPKSKEEG--E 243  
DB 184 QPASTLSQAAYRPOQSOQOTAYSOQRPPPELSDSDSGSAASPMTSKGGOE 243  
OY 244 DLQSLICVARRVPMKEPVLPSSESFTTRDQKITSLDTSTMRAAKPGMEDLVRC 303  
DB 244 DM-----NLSQSNR-----SSLPLDGSIDDLPMGT-EGALSPG----- 277  
OY 304 IQKHAQHEGVSATAKHHHEVLRQGLAFSOIYRFSLSDGTLVAQTKSLINSQTNE 363  
DB 278 -----VSTSGISSSQ-----GEOQNP 293  
OY 364 POLYLSLMLHREQNVCVNPDLTGOTMGKPLNPSSNSPAHQALSGNPGQDMLSSNI 423  
DB 294 AOSPPSPH-----TSPHLPG-INGPSPSPV--GSPASVAGSRGCP-----LS--- 332  
OY 424 NFPIPGKQGMMPKRGPGSGGNHVSQMOATTPQGSNTYALK--MNSPQSSPGMNPQP 482  
DB 333 --PAVPGNQK-PPRPPGSGDSIMHPSMNOSSIADRGYQMRNPQMOYSSP-----QP 384  
OY 483 TSMSPRIRMSPGVAGSRIRPPSGSPAGSLHSPVGCSSSTGNSISYINSSINMALQALSE 542  
DB 385 GSALSPRQ-----PS-----GGQIHTGMG-----SYOONSMSGYPGOG 418  
OY 543 GHGVSLSGLASPLDKMGNLQNSPVNMPPLSKGMSLSDKDCFG-LVGEPESEGTGQAE 601  
DB 419 QYGRQ-GGYPRQPN-----YNALPNANYPSGAMAGGIMPACAGOMGQRPITPYG--- 468  
OY 602 SSCHPGEQKET-----NDPNLPPAVSSERADGOSRLHDSKQTKLLDILLTTSQDM 652  
DB 469 -TLPPGRKSHASMGNRPGYPMANMPPOVSGMCPPPGCMNKKQTQETAVAMHVANSION 527  
OY 653 EPSPLASLSLSTNDSSTGLSPGSGSTHGTSLKEKHKIIHLRLQDSSSYVDL----- 703  
DB 528 RPPGYPMN-----NOGMMGTGPYGGQINS-----MAGMINPGPPYSKMGTMANNSA 576  
OY 704 -----AKLT-----AEATGKDLG--OESSSTAAGSEVITIQEVPSPRK 739  
DB 577 GMAASPENMGIGDYKLPATKMNKADPTPTESKSSSSSTTNEKITIKEYLGEPE 636  
OY 740 KENALLRYLLDKD-----TKDI-GUPEITPKLERLDSKTD 774  
DB 637 RKMVVDRIATFEERKAMGTNLPVGRKPLDLRLYVSVEKEIGLGYVKNKKKRELATN 696  
OY 775 PASNTKLIAMTEKE-----EMSFEPGDQPSSELDNEEILDDIQNSQLPOLFPDT 825  
DB 697 LNVSTSSAASLKKQYIQLYAECKIERGEDPPDI-----FAAADSKKSQ-----FKI 747  
OY 826 RPPGAPAGSVDDQ-----AIINDLMQLTAEENSP--VVP-----VGAOKT 861  
DB 748 QPPSPAGSGMGOPTPOSTSSMAEGDLKPRTPASTPHSIOIPLPQKMSNSNGVIGIDA 807  
OY 862 ALRISQSTFN-----NPRPG-----QLGRLL--NQNLPLDITLQSPTAGP--- 902

DB 808 FNDGSDSTFOKRNKSMTPNPGYQPSMNTSDMMGRMSYEENKD-PYGSNRKAP-GSDPFMSS 865  
OY 903 ---PPIRNSSPYSVLPQCGMNGMIGNQ-----GMLGSSSGMIGN 942  
DB 866 GGGPFGMGDPYSKRAAGP-IGNVAMGRORHYVGGPYDRVTEPGIGEGMMSG- 920  
OY 943 SASRPTMSGEWAPOSSAVRVTCAATTSAMNRPVQGMIRNPAASIPMRSSPOQOROTL 1002  
DB 921 -----ARQPNL-----MPSNDSGM-VSPKPYPOQOQO 948  
OY 1003 OSOVNIGPSELEMMNGGPOYSQQAAPPNOTAPWPESTILPIDQASFASONRQPFSSPDD 1062  
DB 949 OQO-----RHDSYGNQFSTQGTGTPSG-----SPPSQQTMYOQOQOQYKRPMDG 992  
OY 1063 LLCPPAESPDEGALLDQLYLALRNFDEIDRALGIPFLVSQSAVDEQESSQDS 1122  
DB 993 TYGP-PARK--HEG-----EMT-----SVPTSGQOQ--POQ----- 1019  
OY 1123 NIMLBOKAPVFPQOYASQAOQAGSYSPMODPNTFTWGORPSTYATLRMQPRGLPRTGLV 1182  
DB 1020 ---QQLPPAOPQP-ASQOQAAQP--SPQOQVYNOYGNAYPATATATERRRAGCP----- 1068  
OY 1183 QNQ-ENQL--RLQDHLQAQONRQPLM-----NOISVSNV 1216  
DB 1069 QNQPFQFGDRDVSAPPGTNAQNMPPQMMGGPIQASAEVAQGTMMQGRNDMTYANR 1128  
OY 1217 NITLRPVPTQAPINQMLAORRETLNQLRQRMHQOQOYQORTLMMKQGLMPTPSM 1276  
DB 1129 OST--GSAPQGP--AYHGVNRTDEML--HTDORANHEGS-----WPSGTRQPP-- 1171  
OY 1277 VAPSGMPATMSNPRIPQANAOQFPPPPNYGISQODPPTGTATTPQSLMSP---RMAAT 1333  
DB 1172 YGFSAPVPPMTNP--PPSNYQPPPSKQNH-IPQVSSP-----APLPPMENTTSFS 1219  
OY 1334 QSPM---MQOSQANP---AYQAPSDI 1353  
DB 1220 KSPFLHSGKMKQKAGPPVPAISHAPV 1247

RESULT 12  
PASI\_MOUSE STANDARD: PRT: 874 AA.  
AC P97481; 008787; 055046;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 01-MAR-2002 (Rel. 41, Last annotation update)  
DE Endothelial PAS domain protein 1 (EPAS-1) (HIF-1 alpha-like factor)  
DE (MHLF) (HIF-related factor) (HRF).  
GN EPAS1.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Brain;  
RX MEDLINE=97152468; PubMed=9000051;  
RA Tian H., McKnight S.L., Russell D.W.;  
RT "Endothelial PAS domain protein 1 (EPAS1), a transcription factor  
RT selectively expressed in endothelial cells.";  
RN Genes Dev. 11:72-82(1997).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6; TISSUE=Hypothalamus, and Skeletal muscle;  
RX MEDLINE=97272213; PubMed=9113979;  
RA Ema M., Taya S., Yokotani N., Sogawa K., Matsuda Y.,  
RA Fujii-Kuriyama Y.;  
RT "A novel bHLH-PAS factor with close sequence similarity to hypoxia-  
RT inducible factor 1alpha regulates the VEGF expression and is  
RT potentially involved in lung and vascular development.";  
RN Proc. Natl. Acad. Sci. U.S.A. 94:4273-4278(1997).  
RN [3]  
RP SEQUENCE FROM N.A.



```

Db 693 ----ARGPYMSPA-----MIALSNKTLKRGQEYE-----E 720
Oy 820 QLPPTTRGAPAGSVDKAIIINDLMQLTRENSPVTPVCAQKATALRISOSTFNNPRGQLG 879
Db 721 QAFODTSGDDPGTSSHLMMKRMKSLMGCTCLPMP-----D 757
Oy 880 RLPLPONTPLDTLTQSPGAGPFPPIRNSPSVYIPQPMGNQGNQ-----930
Db 758 KITSANMAPDEFTOKSMKGLG--OPLRLPP-----POPSTSSGEMNATGFPPOCYASQ 811
Oy 931 --NLGNSVTGMIGNSASRPTMPSGEMAPDOSAIVYTCATTSAMNRPYOG 978
Db 812 FQDYPGPPGAQKVSVAASRLGSPFEPYLLPELFRYDC-----EVANVPVG 856

RESULT 13
PAS1_HUMAN STANDARD: PRT; 870 AA.
AC 099814; 099630:
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Endothelial PAS domain protein 1 (EPAS-1) (Member of PAS protein 2)
DE (MOP2).
GN EPAS1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97152468; PubMed=9000051;
RA Tian H., McKnight S.L., Russell D.W.;
RT "Endothelial PAS domain protein 1 (EPAS1), a transcription factor
RT selectively expressed in endothelial cells.";
RL Genes Dev. 11:72-82(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Hepatoma;
RX MEDLINE=9079689; PubMed=9079689;
RA Hogenesch J.B., Chan W.K., Jackiw V.H., Brown R.C., Gu Y.-Z.,
RA Priy-Grant M., Perdev G.H., Bradford C.A.;
RT "Characterization of a subset of the basic-helix-loop-helix-PAS
RT superfamily that interacts with components of the dioxin signaling
RT pathway.";
RL J. Biol. Chem. 272:8581-8593(1997).
CC -1- FUNCTION: TRANSCRIPTION FACTOR INVOLVED IN THE INDUCTION OF OXYGEN
CC REGULATED GENES. SPECIFICALLY RECOGNIZES AN 8 BP HYPOXIA RESPONSE
CC ELEMENT (HRE). REGULATES THE VASCULAR ENDOTHELIAL GROWTH FACTOR
CC (VEGF) EXPRESSION AND SEEMS TO BE IMPLICATED IN THE DEVELOPMENT OF
CC BLOOD VESSELS AND THE TUBULAR SYSTEM OF LUNG. MAY ALSO PLAY A ROLE
CC IN THE FORMATION OF THE ENDOTHELIUM THAT GIVES RISE TO THE BLOOD
CC BRAIN BARRIER. POTENT ACTIVATOR OF THE TIE-2 TYROSINE KINASE
CC EXPRESSION.
CC -1- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER
CC BHLH PROTEIN, HETERODIMER WITH THE ARNT PROTEIN.
CC -1- SUBCELLULAR LOCATION: Nuclear (potential).
CC -1- TISSUE SPECIFICITY: EXPRESSED IN MOST TISSUES, WITH HIGHEST LEVELS
CC IN PLACENTA, LUNG AND HEART. SELECTIVELY EXPRESSED IN ENDOTHELIAL
CC CELLS.
CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
CC TRANSCRIPTION FACTORS. HIGHEST TO HIF-1 ALPHA.
CC -1- SIMILARITY: CONTAINS 2 PAS (PER-ARNT-SIM) DIMERIZATION DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 PAS-ASSOCIATED C-TERMINAL (PAC) DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

CC -----
DR EMBL: U81984; AAB41495.1; -
DR EMBL: U51626; AAC51212.1; -
DR TRANSFAC: T02718; -.
DR MIM: 603349; -.
DR InterPro: IPR003015; HLH_MYC.
DR InterPro: IPR001092; HLH_dim.
DR InterPro: IPR001610; PAC.
DR InterPro: IPR000014; PAS.
DR Pfam: PF00785; PAC; 1.
DR Pfam: PF00989; PAC; 2.
DR SMART: SM00353; HLH; 1.
DR SMART: SM00086; PAC; 1.
DR SMART: SM00091; PAS; 2.
DR PROSITE: PS00038; HELIX_LOOP_HELIX; 1.
DR PROSITE: PS50112; PAS; 2.
DR Repeat: DNA-binding; Nuclear protein; Transcription regulation;
DR Activator; Angiogenesis; Developmental protein.
KW DNA_BIND
FT 15 27 BASIC DOMAIN.
FT 28 68 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).
FT 84 154 PAS 1.
FT 230 300 PAS 2.
FT 304 347 PAC.
FT 474 480 POLY-SER.
FT 60 60 A -> E (IN REF. 1).
FT 539 539 G -> D (IN REF. 1).
FT 601 601 R -> H (IN REF. 1).
FT 693 693 N -> D (IN REF. 1).
FT 716 716 K -> E (IN REF. 1).
FT 722 722 P -> L (IN REF. 1).
FT 765 765 L -> F (IN REF. 1).
FT 769 769 S -> P (IN REF. 1).
FT 844 844 R -> C (IN REF. 1).
FT 847 847 K -> N (IN REF. 1).
SQ SEQUENCE 870 AA; 96425 MW; 3DF5B7B13AEC871D CRC64;

```

Query Match 3.6%; Score 278.5; DB 1; Length 870;  
Best Local Similarity 20.1%; Pred. No. 8.9e-06;  
Matches 207; Conservative 140; Mismatches 394; Indels 289; Gaps 44;

```

Oy 10 DSRARETRKRCPCPOLGSPKRNTEKRNRDEENYIEELA-ELIFANPDIONDFKPD 68
Db 6 EKRRSSSEKRE-----KSRDAARCRKSETEFYFLAHLPLPH-----SVSSHLD 52
Oy 69 KCAILKEFYKQIRQIKEDKAAANIDEVOKSDVSSTGGVIDKALGPMLEALDGEFF 128
Db 53 KASIMRLAISFLRTHKLLSSVCSSESEAEAD-----QOMDLYLKALEGFIA 100
Oy 129 VVNEGNAVVESENTQYIRYNOEELMKNSVYSLHVGDTFEVNLPLKSTVNGSGMSG 188
Db 101 VVTQGDMLFLESENISKFEGLTQVELTGHSDFTPHCEIRNL---SLKNSGFG- 156
Oy 189 EPPRRNSHTFNCBMLVKPLPDSEEGHDQEAHQYETFMQCAVAPQSKIKEEG-----242
Db 157 -----KSKDMSTRDPPFMKAKCVTVNRGRVNLKSAFTKVL 193
Oy 243 -----EDLOSCLICVARRVPKREPLV-SSSEFTTRDLOGK 279
Db 194 HCTGVKVVNCPNPNHSLGCVKEPLSLCILICEPIQHSMDIPDLSKTFLSRSMQK 253
Oy 280 ITSLSDTSTRAAMKRGW--EDLVRCIOKFNQHGSESVYAKRHHNHEVLRGLAFSQIT 337
Db 254 FYICDD--RIETELGYHPEELGLRSAYEYFHALDSENNMT--KSHONLCYTGQVVSQY 307
Oy 338 FRSLSDGLTVAQOTKSKILRSQTNBPOLVSLH--MLHREONVCVMNPDLTGOTMGRP- 394
Db 308 RMLAHGCVWLEETGTYIYNPRNLQPOCINCVNVLSIENDDVYFSMDQY-ESLFRPH 366
Oy 395 ---LNPISNSPQAQALCSGNGQDMTSSNINFPINGRKE-----QMGMPDG-----439
Db 367 LMAANSIFDSS-----GKG-AVSEKSNFLFKLKEPEELAQALPAPGDAITSL 414

```



```

QY 762 ITPKLELRDSTKTPASNTKLA-----MKTE-----KEEMSPFGDQPGS--ELD 804
DB 449 SAPNLSSTV-SQIDPSIRRAVALGLPYQVNMPTQPOVOAKNOQPGSGPQGMREKS 507
QY 805 NIEEHLDLQNS--QLPQLEPPTPRGAPAGSVDAKQAIINDLMQLTJENSPTVPYCAQKT 861
DB 508 NMSASPMGVNGVGQVTPSLSDSM-----LHSAINSQNPMMSENAVSPLSGPMT 558
QY 862 AL-----RISQSTFNNRPQGLG-----879
DB 559 AAQPSGTGIRKQWHDITQDLRNHLVHKLVOAIFPTPDPAALKDRRMENLVAYARKVGD 618
QY 880 -----RLTPNQNLPLDITLOS-----PTGAGPPIRRSSPYSTIP 915
DB 619 MYESANNRAEYTHLAEKTYIKQKLEEKRRTRLOKQMLPAGAGMVSANPGRMGOP 678
QY 916 QPMKMGNOGMIGNOGNLSSTGMTIGNSASRPTMPSGEMAPQSSAVRYTCAATTSAMNR 975
DB 679 QPGMTSN-----GLPDPDS--MIRGSVPNQOMP-----704
QY 976 VQGGIRNPASIPMRSSQGOQROTLOSQVYNITPSELEMMNGGPQYSGQAAPPNGTAP 1035
DB 705 -----RITPQSGLNQFG---QMSMAQPIVPRQTPQLQ---734
QY 1036 WPESILPIDQASFASQNRQPGSSPDLLCPHPAESPSDEGALLDQLYLALRNFDGLEE 1095
DB 735 -----HHQQLAQPGALNRP--MGYGRMQQPSNQGOLPQO-----767
QY 1096 IDRALGIPELVYSOASOAVDEQFSSQDSNIMLEOKAPVPRQOYASQAOAAGS--YSPMQ 1152
DB 768 -----TOFPSQGMVNTNPLAPSSGQAPVSOAQOASSSCPVNSPIM 808
QY 1153 DPNHTMGQBSYATLQMRQPRGLRPTGLVQNPQNLQLOLQHLQAOQNQPLMNQISN 1212
DB 809 PP-----GSQGSHTHCPLQPPALH-----QNSPSP-----834
QY 1213 VSNVNLTRPGVPTQAPITNAQMLAQRQREILNQHLRQRMHQOQVOQRTLLMRGQGLNM 1272
DB 835 -----VPSKTRP-----TPH-----HTPSSI-----GAQO 853
QY 1273 TPSSVADSGMATNSNPRIQANA-----QQFPPPNYGISQDPDPGTGATTPOSPLMS 1327
DB 854 PPATTTIPARVPTPRAMPGRPOSQALHPPRRQTPPTTQLPQOVQPSLAPAPSAQDPQOQ 913
QY 1328 PRMAHT-----QSPVMOQSQAAPAYQAPSDING 1355
DB 914 PRSQOSTRAASVTPVAPLPPQAPATPILSQPAVSIEG 949

```

```

CC CC DURING EMBRYOGENESIS AND IN THE ADULT.
CC -1- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER
CC BHLH PROTEIN. HETERODIMER OF SIM1 AND ARNT.
CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
CC TRANSCRIPTION FACTORS.
CC -1- SIMILARITY: CONTAINS 2 PAS (PER-ARNT-SIM) DIMERIZATION DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 PAS-ASSOCIATED C-TERMINAL (PAC) DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb.sib.ch).
CC
CC EMBL: U70212; AAB62395.1; -.
CC InterPro: IPR003015; HLH_Myc.
CC InterPro: IPR001092; HLH_dim.
CC InterPro: IPR001610; PAC.
CC InterPro: IPR000014; PAS.
CC Pfam: PF00785; PAC; 1.
CC Pfam: PF00989; PAS; 2.
CC SMART: SM00353; HLH; 1.
CC SMART: SM00086; PAC; 1.
CC SMART: SM00091; PAS; 2.
CC PROSITE: PS00036; HELIX_LOOP_HELIX; 1.
CC PROSITE: PS50112; PAS; 2.
CC Developmental protein; Neurogenesis; Nuclear protein; Repeat;
CC Transcription regulation; DNA-binding.
CC DNA_BIND 1 13 BASIC DOMAIN.
CC DOMAIN 14 54 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).
CC DOMAIN 77 147 PAS 1.
CC FT 218 288 PAS 2.
CC SQ SEQUENCE 766 AA; 85474 MW; 6D33694BF9A29DD6 CRC64;

```

Query Match 3.6%; Score 273.5; DB 1; Length 766;  
Best Local Similarity 20.9%; Pred. No. 1.3e-05;  
Matches 168; Conservative 112; Mismatches 292; Indels 233; Gaps 36;

```

QY 32 ENTERKRENEENKYTELEAFELIFANFNIDNPNFPDPCALIKETVKQIR-OIKQEKAA 90
DB 6 KNAATFRREKENSEFEYELAKLPL-----ASAITSQVDKASIIIRLTSYIAKRVVFPDEGL 61
QY 91 AANIDEVQKSDVSTGQGVIDKDALGPMLEALDGFVFNVLDEGNVVSERNVQYLRYN 150
DB 62 EAMGHSRTPSLDNVGR-----ELGSHLQTLQDGFIVVAPDKIMVISTASVHLGLS 115
QY 151 QBELMANKSVYSILHVGDTPEVKNILPKRSIVNGSGWSGEPFRNSHTNCHMLYKPLPDS 210
DB 116 QVELTGNISIVEYIHPADHDEMTAVL-----140
QY 211 EEEGHQNEAH--OKYE-----TMOCFAVSOPRSIKEEGEDLQSC-----248
DB 141 --TAHQPYHSHFVQDETERSEFELRMKCYLAKRNAGLCGGKYVHCGYLIKRIQYSLDM 198
QY 249 -----LICVARRVPKKERVPVPS-----SESEFTTRQDLQKITSIDTSTMR 289
DB 199 SPFGQCYQNVGLVAVGHSLP-----PSAVTEIKLHNMFMFRASLDMKILFLDS---R 248
QY 290 AAMKPGWE--DLVRRCLQKFLPAQHEGESVYAKRHHNEVLROGLAFISOIYRFSLSLDTLV 347
DB 249 VAEITGYEPDOLIEKTL--YHNVH-GCDTFLRCAHHLLLVKGVQTVTKYVFFLAKHGGWV 305
QY 348 AAQTKSKIRISQTNPEQVLVSLIMLHREQVNCVANNPDLTQGTQMKPLNPITSSSPAQA 407
DB 306 WVQSTATTIVHNSRSSRPRCIVS-----NYVLTDYEYKILQLS--LDQISAKSPAFSY 356
QY 408 LCSGNPGQDMTLSSNINPPIGPKPEQMGMPGRGSGGGMHVAGMQATT--POGSNV-A 464
DB 357 TSSSTP-----TMTDN-----RKGAHSRLSSSKSKTSRTSPQYSGHNT 395

```

CC -1- FUNCTION: TRANSCRIPTIONAL FACTOR THAT MAY HAVE PLEIOTROPIC EFFECTS





```

Db 606 TKA-----RTQDALAPKQOQOQOQVGTNNQOQOLAVATAQOQOQOQLTAAALQR 658
QY 947 PTMP-----SGEMAPQSSAVRYTCATTSAMNRPVGGMIRNPASISPMRPSQPGQOTL 1002
Db 659 PGAPVYMHNGVQVPPASSV-----STQTAQNSILKAKMKRKOQ--FVYRA-----LATL 706
QY 1003 QSOVYMTI-GPSELEMMNGGPQYSSQ-----GAPNQTAPWPESTILPIDQASFASQNR 1053
Db 707 KTEIGVAGQNKVYGLTLYTQOQOQATNLQOVNAAGNKVYMTSTGPPI-----TLQNG 761
QY 1054 QPRESSDDLCPRPAESPSDECALLDQLYLALRNFDGLEDIDRALGIPELVYSQSAVD 1113
Db 762 QTL-----HAATAAGVADKQOQOQLQFOKQOIIQQOQOMLQOQIAIOMQOQQAAYQ 811
QY 1114 PEQFS-----SODSNIMLEOKAPVPPQYA--SQAMAGQSVSPMDPNPHMGORSPYATL 1168
Db 812 AQOQOQOQOQOQOQVNAQOQOQAAVAAQOQOQOQ-----REQOQOQVAAQOQ 860
QY 1169 RMPQRPGLRPTGLVQNPQNLRLQLHRLAQQRNPQLMNOISVSNVNLTLRPQVPTQA 1228
Db 861 QHQOQALNATQOQLQVAPNPFITSHQOQOQOQQLHNQOQLQOQ-----QAQA 908
QY 1229 PINAQMLAQQRRLNHLQRQMHQOQOQOQVQOQRTLMAR-----GQ-GL 1270
Db 909 QVQAQVAAQOQOQOQ-----QQOQEQOQOQNIQQIVQOQSGATSQOQOQHQSGQLQL 961
QY 1271 NMTPSVAPSGMPATMSNPRIPOANAOQFPPRPVYGISQOQDPGFTGATTPQOS----- 1323
Db 962 SSVPFSVSSSTTPA-----GIATSSALQALASASGALFQAKPCTGSSSTSTSVVITTN 1016
QY 1324 PLM-SPRMAHTQSPMMAQSQANPAPAYQAPSDINGMAQNMGNMFSQOQSPPHGCOQ 1378
Db 1017 QSSPPLTSSSTVASIQQAQOQSAQVHQHQLIS-----ATINGGQQOQOQPGPSLPT 1069
QY 1379 AATSMASNNANINVSATNTGMSNMAMQGLSMSTVTS-----VTSIGLSSMG 1428
Db 1070 TNPILAMTSM-----MNAATGHLSTAPPVYVSTAVTSSPGLVLLSTASSGGGSGIP 1124
QY 1429 PEQVNDPALRG--GNLEPNQLPGMDMKOESDPTRK 1462
Db 1125 ATPKETPPSKGPTATVLPICSPKTPVSGKDCPTTPK 1160

```

```

RESULT 17
HIFA_HUMAN
ID HIFA_HUMAN STANDARD; PRT: 826 AA.
AC Q16655;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypoxia-inducible factor 1 alpha (HIF-1 alpha) (ARNT interacting protein) (Member of PAS protein 1) (MOP1) (HIF1 alpha).
GN HIF1A.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 166-170; 259-289 AND 771-781.
RX MEDLINE=9236340; PubMed=7539918;
RA Wang G.L., Jiang B.-H., Rue E.A., Semenza G.L.;
RT "Hypoxia-inducible factor 1 is a basic-helix-loop-helix-PAS heterodimer regulated by cellular O2 tension.";
RL Proc. Natl. Acad. Sci. U.S.A. 92:5510-5514(1995).
RP [2]
RN SEQUENCE FROM N.A.
RX TISSUE=Hepatoma;
RA MEDLINE=97236817; PubMed=9079689;
RA Hogenesch J.B., Chan W.K., Jackiw V.H., Brown R.C., Gu Y.-Z.,
RA Pray-Grant M., Perdev G.H., Bradfield C.A.;
RT "Characterization of a subset of the basic-helix-loop-helix-PAS superfamily that interacts with components of the dioxin signaling pathway.";

```

```

RL J. Biol. Chem. 272:8581-8593(1997).
RN [3]
RP SEQUENCE FROM N.A.
RA Rupert J.L., Hochachka P.W.;
RT "Hifa sequence in the quechua, a high altitude population.";
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: INVOLVED IN THE INDUCTION OF OXYGEN REGULATED GENES.
CC SPECIFICALLY RECOGNIZES AN 8 BP HYPOXIA RESPONSE ELEMENT (HRE).
CC -1- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER BHLH PROTEIN. HETERODIMER OF AN ALPHA AND A BETA (ARNT) SUBUNITS.
CC THE ALPHA SUBUNIT IS UNIQUE TO HIF-1 WHEREAS HIF-1 BETA (ARNT) CAN DIMERIZE WITH OTHER BHLH-PAS PROTEINS.
CC -1- SUBCELLULAR LOCATION: Nuclear (potential).
CC -1- TISSUE SPECIFICITY: EXPRESSED IN MOST TISSUES. HIGHEST LEVELS IN KIDNEY AND HEART.
CC -1- INDUCTION: UNDER REDUCED OXYGEN TENSION.
CC -1- DOMAIN: BOTH HYPOXIA RESPONSIVENESS AND TRANSACTIVATION CAPABILITY RESIDE WITHIN THE C-TERMINAL PART.
CC -1- PTM: REQUIRES PHOSPHORYLATION FOR DNA-BINDING.
CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF TRANSCRIPTION FACTORS.
CC -1- SIMILARITY: CONTAINS 2 PAS (PER-ARNT-SIM) DIMERIZATION DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 PAS-ASSOCIATED C-TERMINAL (PAC) DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U23431; AAC50152.1; -.
DR EMBL: U29165; AAC51210.1; -.
DR EMBL: AF207601; AAF20139.1; -.
DR EMBL: AF207602; AAF20140.1; -.
DR EMBL: AF208487; AAF20149.1; -.
DR TRANSFAC: T01610; -.
DR MIM: 603348; -.
DR InterPro: IPR003015; HLH_Myc.
DR InterPro: IPR001092; HLH_dim.
DR InterPro: IPR001321; Hypoxindf1A.
DR InterPro: IPR001610; PAC.
DR InterPro: IPR000014; PAS.
DR Pfam: PF00785; PAC; 1.
DR Pfam: PF00989; PAS; 2.
DR PRINTS: PR01080; HYPOXIAIF1A.
DR SMART: SM00353; HLH; 1.
DR SMART: SM00086; PAC; 1.
DR SMART: SM00091; PAS; 2.
DR PROSITE: PS00038; HELIX_LOOP_HELIX; 1.
DR PROSITE: PS50112; PAS; 2.
KW Repeat; DNA-binding; Nuclear protein; Transcription regulation;
KW Activator; Phosphorylation.
FT DNA_BIND 17 30 BASIC DOMAIN.
FT DOMAIN 31 71 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).
FT DOMAIN 85 158 PAS 1.
FT DOMAIN 228 298 PAS 2.
FT DOMAIN 302 345 PAC.
FT DOMAIN 615 621 POLY-THR.
SQ SEQUENCE 826 AA; 92670 MW; ABDAF7DA135BE2D CRC64;

```

Query Match 3.58; Score 269.5; DB 1; Length 826;

Best Local Similarity 19.38; Pred. No. 2.1e-05;

Matches 186; Conservative 150; Mismatches 337; Indels 289; Gaps 41;

```

QY 1 MSGCGENTSPSRAETKRKCEPDGLSPKRNTERKNNREQENKYELELIFANNDI 60
Db 1 MEGAG-GANDKKKISSVRKE-----KSDAARSRKSEVSEYELAHQPLPHNVS 51
QY 61 DNFNFKPKDCAILKEVKAQKQIKQEKKAANIDEVQKSDVSTGCGIDKALGPMML 120

```



FT DOMAIN 1794 1799 POLY-GLN.  
 FT DOMAIN 1856 1863 POLY-GLN.  
 FT DOMAIN 2044 2059 POLY-PRO.  
 FT DOMAIN 2405 2408 POLY-ALA.  
 FT DOMAIN 3216 3220 POLY-PRO.  
 FT DOMAIN 3380 3409 POLY-GLN.  
 FT DOMAIN 3412 3420 POLY-GLN.  
 FT DOMAIN 3534 3550 POLY-GLY.  
 FT DOMAIN 3620 3623 POLY-PRO.  
 FT DOMAIN 3659 3662 POLY-SER.  
 SQ SEQUENCE 3726 AA: 406567 MW: 915ACBE588A72C98 CRC64:

Query Match 3.5%; Score 268.5; DB 1; Length 3726;  
 Best Local Similarity 19.5%; Pred. No. 0.00016;  
 Matches 285; Conservative 170; Mismatches 574; Indels 433; Gaps 65;

Oy 32 RNTERRNREQENKYLE-ELAEILIFANFDIDNFRPKDCAILKETVKQIROIKEDQKAA 90  
 Db 2301 ROKARKNENQEGKDEGERRELTNDRIYRTSNLYOCKKCSLVFORIFDL-IKHOKKLC 2358  
 Oy 91 AANDIVOKSDVSTGCGVIDKDAL-----GPMLEALDGFVVNLEGNVAV 139  
 Db 2359 YKDEDEEGODD-----SQNEDSMAMEILPTSSSCSTPMPSQAYS-----TPASAA 2408  
 Oy 140 SENVTQVLRNQEELMKNKSVSILHVGDHTEFEVKMLPKSIYNGSGWSGEPRRNSHTFN 199  
 Db 2409 NTAPSAFQLQRLAETDELATFNKSKAPASDEK-----FK-----QADPPS----- 2446  
 Oy 200 CRMVLKPLPDESEEGHDNOEAHQKYEEM-QCFVSAQSPKSIKEGEDLOSCLICVARVPM 258  
 Db 2447 ---AOPNTOEKGOPKPRPMOQOLEQOKTNAPOPKLPAPASL-----PQPPPO 2495  
 Oy 259 KERPLPSESESTTODLOG-KITSLDYST-----MRAAMPGEEDLYRCIOKFAH 311  
 Db 2496 APPPCPLPQSSPSPQSLSHLPLKPLHTSTPQOLANLPOLLPYOCDCKLAFPESEHMO 2555  
 Oy 312 EGESVYAKRHHHEVLRGOIAFSQIYRFLSD--GTLVAOAKSKLIRSQTNEDPOLVTS 369  
 Db 2556 EHQQLHFLSAQNOFIHPQFLRLSLMPFLPSPNPLASQLSGLAI-----POIPAS 2608  
 Oy 370 LHMILREONVCVMDPLTGOTMGKPLNPILSSNSPAHQALCSG--NPGQDMTSLSNINPFI 427  
 Db 2609 ---SATSPSTPTSTMTLTKLEKASAPGENDSGTGEPEORDKRLRTTIT--- 2658  
 Oy 428 NCKPKOMGMKGRFGSGG-----MNHVS---GMQATTPOGSNTYALKMSPQSSGSG--- 476  
 Db 2659 ---PQOLELYOKYLLDSNPTKRLMDHIAHEVGLKRRVQO---VWFQNTRARERKQOPR 2711  
 Oy 477 -MNPQOPTSMLSPRHRMSR-----GVAGSPRIIPPSQFSPA----- 510  
 Db 2712 AVGPQO-----AHRRCRPFCAKAKTALFAHIRSRMHAKRAGVYMLTISAMILDCD 2764  
 Oy 511 GSLHSPVGCSTGNSHSTYNSLNAOLAISGHCVSILG-----SSLASP-DLKM 559  
 Db 2765 GGIQKKGDIIFDCTSFSSHLPSS-----SDQGVPLSPVSKTMELSPTLLSPSSIKV 2816  
 Oy 560 GNLOK-----SPVMNMPPLSKMGSIDSKOCFGLXGPESEIT---GQAESCHPGEK 610  
 Db 2817 EGIEDFESPMSSVNLN---FDQTKLNDCCSVNTAITDTTDTDEGNADNDSATGIAIT 2872  
 Oy 611 ETND--PN--LPPAVSERADGOSRL-----HDSKQOTKLULLTTKSPQ 651  
 Db 2873 ETKSAPNEGILKAMAMMEYEDRLSGVSPASFSKEDONECTVUYSE-TSSLADP 2931  
 Oy 652 MEPSLASSLSDTNKSDTSGSLPGSGSTGTSLKEKHKILHRLLODSSSPVDLAKLTAET 711  
 Db 2932 CSPSPGASGASGKSGDG--GDRPGOKRFRQMTNQLKYLKSCFNDYRFP---TMECEVL 2987  
 Oy 712 GMDLSQESSSTRPGESEVITIKQEPVSPKKKENALLKYLKDKDTKDIGLPEIT---PKLE- 767  
 Db 2988 GNDIG-----LPRVVOVWFQNMARAKKSKL-----SMAKHFGIMQTSYEGKTEC 3034

Oy 768 -----RLDSKTDPAENFKLIAMKTEKEMSEPEGDPSSELDNLEIIDLQNSQL 818  
 Db 3035 TLGKITYSARLSVRHIFSQHISKV-----DTISQDKKEKEYPDPAVRQL 3083  
 Oy 819 POLFPDTRGADGASVCKOAILINDLMQLTAEKSPVTPVGAQXTALRISOSTNNRPGOL 878  
 Db 3084 -----MAQOELDRIKKANENVGLAAO-----QCGMEDN----- 3111  
 Oy 879 GRLLPNQNLPLITITQSPGACGPPPIRNSPSYVIRPGMGNGMGNGMGNGMGNGMGNG 938  
 Db 3112 ---APLOALNLPIT---YPLQGIIP---VLLPGL----- 3137  
 Oy 939 MIGNSARPEPMPSGEMAPSSAVRVYCAATTSAMNRPVOCGIMRN-----PAASIPMR 991  
 Db 3138 -----NRSLSLG-----FTRPATLTALSKPKNLMGLPSTTVSPGLTSLPRK 3180  
 Oy 992 PSSQPGQROTLOSQVANNIGPSELNNMGSPQYSSQOAPPNQAPWDESI--LPIDQASFA 1049  
 Db 3181 PSSASLSPPTPAQATWAMAPQP-----PQOPQOPQPPVQOPPPPPAQOIPAPQITPQ 3233  
 Oy 1050 SQNRQPFSSPPDLCRHPAESPDEGALLDQVLYALANPGLGEIDALGIPE----- 1104  
 Db 3234 QQRKDKGEGKKEKRAHKGKEP-----LVPKKEGE 3267  
 Oy 1105 -----LVSSQAVDPEQFS-----SODSNIMLEOK-----APVFPQOYASQA 1141  
 Db 3268 APPAGTGITSAFLPMMEVAVDPAQLOALQALTSPTALTALTSQFLPYFPGSPYYAPQI 3327  
 Oy 1142 QMA-QGYSYPMODPNEFTMGORPSYATLRMDRPGILRPGLYVONQRLQLQHRQLAQ 1200  
 Db 3328 PALQSGY---LQPMWGMGLFP--YSPALSRPLMGISPSGLIQ-QYQYQOSLQEAIOQO 3382  
 Oy 1201 QNRQPLMNOISVSNVNLRLRCGVPTQAPINQMAQORRELLNQLRRORHQQOQYQO 1260  
 Db 3383 QQQQQQQQQQ-----QQQQQQQO-----LQQQQQQQQQQQYQO 3413  
 Oy 1261 RFLMRGQGLNMTTP---SWAPSGMPATWS-----NPRIPQANAOQPPPPYNGISQOPD 1312  
 Db 3414 QQQQQQQQPKASQTPVPOGAASPDKPAKESPRPEQKNVRELSPLLPKPP-----EEPE 3468  
 Oy 1313 PGFTGATTPQ--SPLMSPRMAH 1332  
 Db 3469 AESKSASADSLCDPFTVPRVQY 3490

RESULT 19  
 BMAL\_HUMAN  
 ID BMAL\_HUMAN STANDARD; PRT; 583 AA.  
 AC 000327; 000313; 000314; 000315; 000316; 000317; 099631; 099649;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE BMAL1 protein (Brain and muscle ARNT-like 1) (Member of PAS protein 3)  
 DE (MOP3) (BHLH-PAS protein JAP3).  
 GN BMAL1 OR ARNTL.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxId=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING (ISOFORMS BMAL1 A-F).  
 RC TISSUE=Brain;  
 RX MEDLINE=97289529; PubMed=9144434;  
 RA Ikeda M., Nomura M.;  
 RT "cDNA Cloning and tissue-specific expression of a novel basic  
 RT helix-loop-helix/PAS protein (BMAL1) and identification of  
 RT alternatively spliced variants with alternative translation  
 RT initiation site usage".  
 RL Biochem. Biophys. Res. Commun. 233:258-264(1997).  
 RP [2]  
 RP SEQUENCE FROM N.A. (ISOFORM MOP3).  
 RP TISSUE=Fetal brain;  
 RX MEDLINE=97236817; PubMed=9079689;

RA Hogenesch J.B., Chan W.K., Jackiw V.H., Brown R.C., Gu Y.-Z.,  
 RA Pray-Grant M., Perlew G.H., Bradfield C.A.;  
 RT "Characterization of a subset of the basic-helix-loop-helix-PAS  
 RT superfamily that interacts with components of the diroxin signaling  
 RT pathway";  
 RL J. Biol. Chem. 272:8581-8593(1997).  
 RN  
 RP [3]  
 RP SEQUENCE FROM N.A. (ISOFORM BMALIB).  
 RA Tian H., Russell D.W., McKnight S.L.;  
 RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.  
 RN  
 RP INTERACTION WITH CLOCK.  
 RX MEDLINE=98279137; PubMed=9616112;  
 RA Gekakis N., Staknis D., Nguyen H.B., Davis F.C., Wilsbacher L.D.,  
 RA King D.P., Takahashi J.S., Weitz C.J.;  
 RT "Role of the clock protein in the mammalian circadian mechanism";  
 RL Science 280:1564-1569(1998).  
 CC  
 CC -1- FUNCTION: CLOCK-BMALI HETERODIMERS BIND TO AN E-BOX ELEMENT (3'-  
 CC CACGG-5'), THEREBY ACTIVATING TRANSCRIPTION OF PER1, AND POSSIBLY  
 CC OF OTHER CIRCADIAN CLOCK PROTEINS.  
 CC -1- SUBUNIT: EFFICIENT DNA HETERODIMERIZATION WITH ANOTHER  
 CC BHLH PROTEIN. FORMS AN HETERODIMER WITH CLOCK. INTERACTS WITH  
 CC HSP90; WITH AHR IN VITRO, BUT NOT IN VIVO.  
 CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).  
 CC -1- ALTERNATIVE PRODUCTS: AT LEAST 7 ISOFORMS; BMALIA (SHOWN HERE),  
 CC BMALIB/AP3, BMALIC, BMALID, BMALIE, BMALIF AND MOP3; ARE PRODUCED  
 CC BY ALTERNATIVE SPLICING..  
 CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN THE ADULT BRAIN, SKELETAL  
 CC MUSCLE AND HEART.  
 CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF  
 CC TRANSCRIPTION FACTORS. HOMOLOG OF DROSOPHILA CYCLE PROTEIN.  
 CC -1- SIMILARITY: CONTAINS 2 PAS (PER-ARNT-SIM) DIMERIZATION DOMAINS.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC  
 CC EMBL; D89722; BAI19968.1; -;  
 DR EMBL; AB000812; BAI19935.1; -;  
 DR EMBL; AB000813; BAI19936.1; -;  
 DR EMBL; AB000814; BAI19937.1; -;  
 DR EMBL; AB000815; BAI19938.1; -;  
 DR EMBL; AB000816; BAI19939.1; -;  
 DR EMBL; U51627; AAC51213.1; -;  
 DR EMBL; U60415; AAB37248.1; -;  
 DR MIM; 602550; -;  
 DR InterPro: IPR003015; HLH\_Myc.  
 DR InterPro: IPR001092; HLH\_dim.  
 DR InterPro: IPR001067; NucleusLocator.  
 DR InterPro: IPR001610; PAC.  
 DR InterPro: IPR000014; PAS.  
 DR Pfam; PF00010; HLH; 1.  
 DR Pfam; PF00785; PAC; 1.  
 DR Pfam; PF00989; PAS; 2.  
 DR PRINTS; PR00785; NCTRSLCLOCAT.  
 DR SMART; SM00353; HLH; 1.  
 DR SMART; SM00086; PAC; 1.  
 DR SMART; SM00091; PAS; 2.  
 DR PROSITE; PS00038; HELIX\_LOOP\_HELIX; 1.  
 DR PROSITE; PS50112; PAS; 2.  
 KM Repeat: DNA-binding; Nuclear protein; Transcription regulation;  
 KM Alternative splicing.  
 FT DNA\_BIND 30 42 BASIC DOMAIN.  
 FT DOMAIN 43 83 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).  
 FT DOMAIN 100 172 PAS 1.  
 FT DOMAIN 283 353 PAS 2.  
 FT DOMAIN 358 401 PAC.  
 FT VARSPLIC 1 4 MINI -> MADQRMDISITSDPMSGPTDLSLSLGTSGV

FT VARSPLIC 1 16  
 FT VARSPLIC 181 181  
 FT VARSPLIC 182 583  
 FT VARSPLIC 231 348  
 FT VARSPLIC 235 258  
 FT VARSPLIC 259 583  
 FT VARSPLIC 400 483  
 FT VARSPLIC 484 583  
 FT CONFLICT 26 26  
 FT CONFLICT 80 80  
 FT CONFLICT 130 130  
 FT CONFLICT 216 216  
 FT CONFLICT 221 221  
 FT CONFLICT 375 375  
 FT CONFLICT 470 471  
 SQ SEQUENCE 583 AA; 64206 MW; 2AAB7EBBA471119 CRC64;  
 Query Match 3.5%; Score 267; DB 1; Length 583;  
 Best Local Similarity 23.3%; Pred. No. 1,8e-05;  
 Matches 146; Conservative 69; Mismatches 245; Indels 146; Gaps 29;  
 QY 1 MSGNGENTSDP-----SRAETRRKKECPDOLGSPKRNTERKREQENKYEELAEI 52  
 DB 4 IESMDTQDKDPHGLETEHQGRKRNAREASOI-----EKRRDKMNSFTDELASL 55  
 QY 53 IFANFNIDNENFEPDOKALIKETVQIRQIK-EQKAAANIDENVQSVSTGGQVID 111  
 DB 56 V-----PTCNMMSRKLDLTVLRNAVQHMKTGATNPYETANKKPFPLSD----- 101  
 QY 112 KDALGPMILBALDGFEEVNVILE-GNVVFESENTOYLRYNOEELMKNKSVSILHVGDTTE 170  
 DB 102 -DELKHLILAAAGFLFVVGCDGKILFSESVFKILNYSQNDLIGSLDFDIYHAPDIAX 160  
 QY 171 FVKNLPRK-----IVNGSWSGEPFRNRSHTEPCRMILV-KPLPD 209  
 DB 161 -VKEQLSSSDPTAPRERLIDAKTGLPVKTDITTPGSRILSGARR---SFCRMKCNRPYSK 216  
 QY 210 SEEGCHDNOEAHQYEMQFAVSQPKSIK-----DEGDIQSLICIVA 253  
 DB 217 VEDKDPSTCSKRRADKKSCTIHTSTGYLKSMPPTKMGLEDNEPDNEGCL-SCLVAIG 275  
 QY 254 RR---VPMK-ERPVLPSSSEFTTRDLODKITSLD-TSTMRAAMKPGWEDLVRCIOKF 307  
 DB 276 RLMSHVVPQPVNGEIRKMSMEYVSRNAIDGKPFVFDORATAILALYLP-QLLGCTSCYEY 334  
 QY 308 HAQHEGESVSYAKRHNEVL--ROGLAFSOIYRPSLSDGLVAAQTKSKILRSQTNPEQ 365  
 DB 335 HQ---DDIGHLAECHECHROVLOTREKIT-TNCFYFKIKDGSFETILRSWFSEFMMPWTEVE 389  
 QY 366 LVISLHMLHREQVYCVNPNPLDTGOTMCKPLNPSSNSPAAQALCSGPGQDMTLSSINF 425  
 DB 390 YIVS-----TNTVVLANYLEGGDPTFPOLTAAP-----HSMDSMLPSDEGG 430  
 QY 426 PINGPKEQMGMPMGREGSGCMNHVSG---MQATTPGGSYVALKMNPSOSSPGMNPQOP 482  
 DB 431 PKRHTPVLPSPGTRGAKIGKIMLAELMEIHLIRIS-----SPSSGSSPLNI 481  
 QY 483 TSMLSPRHRMSPG-----VAGSPRIIPSPQSPAGSLHSPGVCSSTGNSHSYNSLINAL 537  
 DB 482 TS--TPPDASSPGKILNGLGTPDIPSSGLSGQAQENP-----GYPYSDS--- 527  
 QY 538 QALSEGHGVS-----GSSLASPD 556  
 DB 528 SILGENPHIGIDMINDOGSSPSND 553



```

QY 801 -SELD-NLEETLDDNSQLPQLPPTRPCAPAGSVKQAI-----NDLM 844
DB 1507 DSEVSSASAGLEEQSOAQPOAPQOT-----SDSVKKEPVLAQPAVNSGTAASSTSLV 1561
QY 845 QLAENSVPVP-VGAOKTALRISOSTFNPNRPGQLRLNQNPLDITITQSPGAGPF- 902
DB 1562 ALSAETPATGTPDARTETAVPPASSFS--VPGQTA-----VIAAATSSACPVA 1608
QY 903 -----PPIRNSPYSVLPQP-----GMNGOMIGNOGNLNSSTG---MIGNSASRP 947
DB 1609 VETSSPTPIASTSTIVAPGSAEAAAGVTGSSSVFAQPPAASSSSAFNQLNNTAT-- 1666
QY 948 TMSGEPAPSSAVRTCAATTSAMNRPVQGMIRNPASIPMRPS-----OPG 997
DB 1667 -----APSATFVGQVAASSTASLEFGQOTGSTASTAATAAPQVSSGFSFSPAFGTAPG 1719
QY 998 Q-ROTLQSOVMNIGPSLE-----MNGGPOYSOOQA--PNNQAPMESTLPIDAQSFAS 1050
DB 1720 VFGQTFPGQASVFGQSSASAVSFSSQPGFSVPAPGAPASSSTPTSTGSGVFGAASSTIS 1779
QY 1051 QNRQ-FPGSSPDDLCPHAPAESPSDEGALLDQLYL-ALBNFDGLEIDRALGPELVSQ 1108
DB 1780 SSSSFSFGQS-----SPNTGGGLFGQSNAPAFQSGFQGGSGVFGTSAATY 1827
QY 1109 SQAV-----DPROFSSQDSNIMLEQAP-----VFPQOYSAQOMAGSYSPMODPNEHT 1158
DB 1828 TAATSGFSGFCOASGFSNGSVFGQAASTGTVFGQSSSSSGSVSGNTGCGGFFS 1887
QY 1159 -MGORPSYATLRMPQRGLPRTGLVQNPQNLRLQLQHLQAQONRPLMN-----Q 1209
DB 1888 GLGGRS-----QDAANKPFSASGCGFSTA 1914
QY 1210 ISNVSNVNLTRPGVPTQABINQMLAQROREILNHLROROMHOQOOVOORTLMRGOG 1269
DB 1915 TSMTS--NLFGNSGAKTFEGFASSFE-----QKPTGTFSSGG 1951
QY 1270 LNMTPSVAVASGMPATIRSNRIPQANAOFPFPNNICISQDPDGFATTPQSLMSRP 1329
DB 1952 ----GSVASQGFSGSSPNTGFGAAPVFGSPTEFGS---PGFGVAPFAGSA----- 1997
QY 1330 MAHTQSPMQSOANPYAQPASDINGMAOGMGSNMFSSQSP-----HFGQANTSM 1383
DB 1998 -----PAFTIP-----LGSTGG-KVFGECTAAASGSGGFGSSMT-- 2032
QY 1384 YSNMNMNINSMATNTGSMSSMNQMT-QQIS-MTSVTSVSTGLSSMG 1428
DB 2033 -----TSFGTLASQNAPTFGLSLQOTSQSGFQSSGFGFG 2067

RESULT 21
SIML_MOUSE
ID SIML_MOUSE STANDARD; PRT; 765 AA.
AC 061045; P70183;
DT 01-NOV-1997 (Rel. 35, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Single-minded homolog 1 (Sim1).
GN SIM1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RC SEQUENCE FROM N.A.
RP STRAIN=SWISS WEBSTER;
RX MEDLINE=97020303; PubMed=8812055;
RA Fan C.-M., Kuwana E., Bulfone A., Fletcher C.F., Copeland N.G.,
RA Jenkins N.A., Crews S., Martinez S., Puellas L., Rubenstein J.L.,
RA Tessier-Lavigne M.;
RT "Expression patterns of two murine homologs of Drosophila
RT single-minded suggest possible roles in embryonic patterning and in
RT the pathogenesis of Down syndrome.";
RL Mol. Cell. Neurosci. 7:1-16(1996).

```

```

RN [2]
RP ERRATUM.
RX MEDLINE=97029422; PubMed=8875433;
RA Fan C.-M., Kuwana E., Bulfone A., Fletcher C.F., Copeland N.G.,
RA Jenkins N.A., Crews S., Martinez S., Puellas L., Rubenstein J.L.,
RA Tessier-Lavigne M.;
RL Mol. Cell. Neurosci. 7:519-519(1996).
RN [3]
RX REVISIONS TO C-TERMINUS.
RP MEDLINE=97343329; PubMed=9199934;
RA Fan C.-M.;
RL Unpublished results, cited by:
RL Cirast R., Scott H.S., Chen H., Kudoh J., Rossier C., Minoshima S.,
RL Wang Y., Shimizu N., Antonarakis S.E.;
RL Genome Res. 7:615-624(1997).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J;
RX MEDLINE=96413339; PubMed=8927054;
RA Ema M., Morita M., Ikawa S., Tanaka M., Matsuda Y., Gotoh O.,
RA Saitoh Y., Fujii H., Hamada H., Kikuchi Y., Fujii-Kuriyama Y.;
RT "Two new members of the murine Sim gene family are transcriptional
RT repressors and show different expression patterns during mouse
RT embryogenesis.";
RL Mol. Cell. Biol. 16:5865-5875(1996).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SV;
RA Hosoya T.;
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SUBUNIT.
RX MEDLINE=97172525; PubMed=9020169;
RA Probst M.R., Fan C.M., Tessier-Lavigne M., Hankinson O.;
RT "Two murine homologs of the Drosophila single-minded protein that
RT interact with the mouse aryl hydrocarbon receptor nuclear
RT translocator protein.";
RL J. Biol. Chem. 272:4451-4457(1997).
CC -1- FUNCTION: TRANSCRIPTIONAL FACTOR THAT MAY HAVE PLEIOTROPIC EFFECTS
CC DURING EMBRYOGENESIS AND IN THE ADULT.
CC -1- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER
CC BHLH PROTEIN. HETERO-DIMER OF SIM1 AND ARNT.
CC -1- TISSUE SPECIFICITY: DETECTED IN LONG, SKELETAL MUSCLE AND KIDNEY.
CC DURING FETAL DEVELOPMENT IT IS FOUND IN THE CNS, DEVELOPING
CC KIDNEY, MESODERMAL AND ENDODERMAL TISSUES, INCLUDING DEVELOPING
CC SOMITES, MESONEPHRIC DUCT, AND FOREGUT.
CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
CC TRANSCRIPTION FACTORS.
CC -1- SIMILARITY: CONTAINS 2 PAS (PER-ARNT-SIM) DIMERIZATION DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 PAS-ASSOCIATED C-TERMINAL (PAC) DOMAIN.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: U40575; AAA91201.1; ALT_SEQ.
DR EMBL: D79209; BAA11467.1; -.
DR EMBL: AB013491; BAA28270.1; -.
DR EMBL: AB013484; BAA28270.1; JOINED.
DR EMBL: AB013485; BAA28270.1; JOINED.
DR EMBL: AB013486; BAA28270.1; JOINED.
DR EMBL: AB013487; BAA28270.1; JOINED.
DR EMBL: AB013488; BAA28270.1; JOINED.
DR EMBL: AB013489; BAA28270.1; JOINED.
DR EMBL: AB013490; BAA28270.1; JOINED.
DR MGD: MGI:98306; Sim1.
DR InterPro: IPR003015; HLH_Myc.
DR InterPro: IPR001092; HLH_dim.
DR InterPro: IPR001067; NucleinsLocator.

```

DR InterPro: IPR001610; PAC.  
 DR InterPro: IPR000014; PAS.  
 DR Pfam: PF00785; PAC; 1.  
 DR Pfam: PF00989; PAS; 2.  
 DR PRINTS: PR00785; NCTRNLOCATR.  
 DR SMART: SM00353; HLH; 1.  
 DR SMART: SM00086; PAC; 1.  
 DR SMART: SM00091; PAS; 2.  
 DR PROSITE: PS00036; HELIX\_LOOP\_HELIX; 1.  
 DR PROSITE: PS00112; PAS; 2.  
 DR Developmental protein; Neurogenesis; Nuclear protein; Repeat;  
 KW Transcription regulation; DNA-binding.  
 FT DNA\_BIND 1 13 BASIC DOMAIN.  
 FT DOMAIN 14 54 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).  
 FT DOMAIN 77 147 PAS 1.  
 FT DOMAIN 218 288 PAS 2.  
 FT CONFLICT 133 133 H->L (IN REF. 1).  
 FT CONFLICT 176 176 MISSING (IN REF. 1).  
 FT CONFLICT 322 322 P->R (IN REF. 1).  
 FT CONFLICT 480 480 A->P (IN REF. 1).  
 FT CONFLICT 537 537 D->S (IN REF. 1).  
 SO SEQUENCE 765 AA; 85540 MW; B1A7FDAB578CD17 CRC64;

Query Match 3.5%; Score 266; DB 1; Length 765;  
 Best local similarity 20.2%; Pred. No. 2.8e-05;  
 Matches 162; Conservative 110; Mismatches 302; Indels 228; Gaps 33;

QY 32 RTEKRNREQENKYLEELAELEIFANFNIDNENFR-----DKCAILEKTVQIR 81  
 DB 6 KNAATREKENSEFEYELAKLL-----PLPSAIRSOLDKASTIRLTYSYKM 52  
 QY 82 QIKDEKAAANIDEVQSDVSSTGQVIDKDALSPMLIALDGFPPVYNLEGNVVFSE 141  
 DB 53 RVFPEGLGELAMGHTSRTPSPLDNVR-----ELGSHLQTLTGDFIVVAPGKIMYISE 106  
 QY 142 NTVQRLRYNOELRNKSVSYILHVDHTEFVKNLLPKSIVNGSGMSGPPRNSTFNCR 201  
 DB 107 TASYHLGISOVELLTGNSITYEIHPRADHDEMTAVL----- 140  
 QY 202 MLVRLPDPSEEGHNDQEN--QKYE-----TMOCFANVSQPKIKERGEDLQSC---- 248  
 DB 141 -----TAHQPYSHFEVQEYIEERSFELRMKCVLAKNAGILTCGGYKVIHCQYL 189  
 QY 249 -----LTCVARYPMKRPVLP-----SESFTRDLOQKI 280  
 DB 190 KIROYSIDMSPFDGQYQNVGLVAVGHSLP-----PSAVTEIKLHNSNFMFRASIDMKL 242  
 QY 281 TSLDTSTMRAMKPGME--DIVRRCIOKFHAOGESESVYAKRNHHEVLRQGLAFSOIYR 338  
 DB 243 IFLDS--RVAEELTGERODLIEKTL--YHNHY-GCDFPHLCANHLLLVKQVYTKYR 296  
 QY 339 FSLSDGLTVAQTKSKLIRSQTTPNEPOLVLSLHMLHREQNCVMMPDILGQTMGRPLNDI 398  
 DB 297 FLAKGGGMMWQSVATYIHNSSRPHCIYV-----NYVLDTPEYKGLDS--LDQI 347  
 QY 399 SSNSNAHOLCSGNPCQDWTLSNINFLNGKPEQMGMPGFGSGGNHVSQMATI-- 457  
 DB 348 SASKTFEYTSSTP-----TISDN-----KKGAKSRSLSSSKSRKSP 386  
 QY 458 -PGGSNY--ALKMNSPQSPGKNNPGQPT--MLSPRHRMSGVAGSPRIIPSPSPAS 512  
 DB 387 FYQISGFITERSESDHDSOMGSPPLTDIASPOLDBER--PGSQHELSCAYRQPPDRSS 443  
 QY 513 L-----HSPVGVCSSTGNHSHYTNSSINA-----LQALSRGHVSLGSSLASPD 557  
 DB 444 LCYGFALDHSRL-----VEDRHFHTQACBGRCEAGRYFLGAPRGDRDPMWSRAALPLT 498  
 QY 558 KMGNTLQNSPVMNPPPLSKMSGLSDKDFGLYGERSEGTGGAESSCHGEQKET-----N 613  
 DB 499 KASPSREAVENSMPIHSIRIHRIR--GHMDEDSVSSPPPGSGSESGDRYRTEYOXON 555  
 QY 614 DPNLPPAVSSERADGQSRLLHDSKQTKILLQLLTTSKSDQNEPPLASSLSDTKKDSGSLP 673

DB 556 SPHEPSKIETTLIRATQOMKKEENR---LQLKAPPDOL-----ASIN 595  
 QY 674 GSGSTHGTSLSAKKH-----KILHRLDSSSPVDLAKITLAFATKDS-----QESSST 722  
 DB 596 GAKKHSILCFANYQOPPTPGGEVCHSSALASTSPCDHIQ---OREGKMLSPHENDYDINSPT 652  
 QY 723 A-----PGSEVITIKOEPPSPK 738  
 DB 653 ALSRISSPSSDRITKSSLLIAK 674

RESULT 22  
 BUN2\_DROME STANDARD; PRT; 1211 AA.  
 AC 024523; Q9VK78; Q9VK79;  
 DT 01-MAR-2002 (Rel. 41, Created)  
 DT 01-MAR-2002 (Rel. 41, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Bunched protein, class 2/class 3 isoforms (Shortsighted protein).  
 GN BUN OR SHS OR CG5461.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A. (CLASS 2 ISOFORM), SUBCELLULAR LOCATION, AND  
 RP FUNCTION.  
 RC TISSUE=Eye-antennal disk;  
 RX MEDLINE=96038094; PubMed=755710;  
 RX Treisman J.E., Lai Z.-C., Rubin G.M.;  
 RT "Shortsighted acts in the decapentaplegic pathway in Drosophila eye  
 RT development and has homology to a mouse TGF-beta-responsive gene";  
 RL Development 121:2835-2845(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A. (CLASS 2 AND CLASS 3 ISOFORMS).  
 RC STRAIN=BERKELEY;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RX Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RX Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RX George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RX Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RX Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
 RX Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RX Abril J.F., Aghayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RX Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RX Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,  
 RX Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,  
 RX Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RX Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RX de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RX Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RX Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,  
 RX Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RX Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RX Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,  
 RX Houslin D., Houston K.A., Howland T.J., Wei M.-H., Ibeagwam C.,  
 RX Jatalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RX Kimmel B.E., Kodira C., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RX Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RX Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RX Merulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RX Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RX Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pauleb J.M.,  
 RX Palazozo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RX Reibert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RX Shue B.C., Siden-Kimms I., Simpson M., Skupski M.P., Smith T.,  
 RX Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RX Svitskas R., Tector C., Turner C., Venter E., Wang A.H., Wang X.,  
 RX Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,  
 RX Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RX Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,





DB 898 SYAAPQALPTLQASAPSTIADPQOLMVPQOQOQOHOE-----EQQ 939

QY 1371 SPHPFGQ-----AN-TSMYSNNMNNINVSMTATGMSNMQTOISMT 1414

DB 940 QQPQOQOQOPLPPANTASASANNLNL---TNTNVATGEATNTNLTTL 985

RESULT 23

AC ARN2\_HUMAN STANDARD: PRT: 706 AA.

AC 09H8Z2: 015024; (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Aryl hydrocarbon receptor nuclear translocator 2 (ARNT protein 2).

GN ARNT2 OR KIAA0307.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RA Barrow L.L., Wines M.E., Romitti P.A., Holener B.C., Murray J.C.;

RT "Aryl-hydrocarbon receptor nuclear translocator 2 (ARNT2): structure, gene mapping, and candidate evaluation for human orofacial clefts.";

RL Submitted (SEP-1999) to the EMBL/Genbank/DBJ databases.

RL [2]

RP SEQUENCE FROM N.A.

RC TISSUE=Brain.

RX MEDLINE=9734984; PubMed=9205841;

RA Nagase T., Ishikawa K.-I., Nakajima D., Ohira M., Seki N., Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;

RT "Prediction of the coding sequences of unidentified human genes. VII. The complete sequences of 100 new cDNA clones from brain which can code for large proteins in vitro.";

RL DNA Res. 4:141-150(1997).

CC -1- FUNCTION: SPECIFICALLY RECOGNIZES THE XENOBIOTIC RESPONSE ELEMENT (XRE).

CC -1- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER HIGH PROTEIN. HETERODIMER WITH THE ARYL HYDROCARBON RECEPTOR (AHR) OR THE SIM1 PROTEIN (By similarity).

CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).

CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF TRANSCRIPTION FACTORS. HIGHEST TO THE ARNT PROTEINS.

CC -1- SIMILARITY: CONTAINS 2 PAS (PER-ARNT-SIM) DIMERIZATION DOMAINS.

CC -1- SIMILARITY: CONTAINS 1 PAS-ASSOCIATED C-TERMINAL (PAC) DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL: AF185610; AAG15310.1; -

DR EMBL: AF185593; AAG15310.1; JOINED.

DR EMBL: AF185594; AAG15310.1; JOINED.

DR EMBL: AF185595; AAG15310.1; JOINED.

DR EMBL: AF185596; AAG15310.1; JOINED.

DR EMBL: AF185597; AAG15310.1; JOINED.

DR EMBL: AF185598; AAG15310.1; JOINED.

DR EMBL: AF185599; AAG15310.1; JOINED.

DR EMBL: AF185600; AAG15310.1; JOINED.

DR EMBL: AF185601; AAG15310.1; JOINED.

DR EMBL: AF185602; AAG15310.1; JOINED.

DR EMBL: AF185603; AAG15310.1; JOINED.

DR EMBL: AF185604; AAG15310.1; JOINED.

DR EMBL: AF185605; AAG15310.1; JOINED.

DR EMBL: AF185606; AAG15310.1; JOINED.

DR EMBL: AF185607; AAG15310.1; JOINED.

DR EMBL: AF185608; AAG15310.1; JOINED.

DR EMBL: AF185609; AAG15310.1; JOINED.

DR EMBL: AB002305; BAA20766.1; -

DR MIM: 606036; -

DR InterPro: IPR001092; HLH\_dlm.

DR InterPro: IPR001067; Nucleinslocator.

DR InterPro: IPR001610; PAC.

DR InterPro: IPR000014; PAS.

DR Pfam: PF00010; HLH; 1.

DR Pfam: PF00989; PAS; 2.

DR PRINTS: PR00785; NCTRNSLOCATR.

DR SMART: SM00353; HLH; 1.

DR SMART: SM00086; PAC; 1.

DR SMART: SM00091; PAS; 2.

DR PROSITE: PS00038; HELIX\_LOOP\_HELIX; 1.

DR PROSITE: PS50112; PAS; 2.

KW DNA-binding; Nuclear protein; Transcription regulation; Repeat.

FT DNA\_BIND 53 65

FT DOMAIN 66 106 BASIC DOMAIN.

FT DOMAIN 123 198 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).

FT DOMAIN 312 382 PAS 1.

FT DOMAIN 387 430 PAS 2.

FT DOMAIN 62 65 PAC.

FT DOMAIN 429 434 POLY-ARG.

FT DOMAIN 541 544 POLY-SER.

FT CONFLICT 64 64 R -> K (IN REF. 2).

FT CONFLICT 668 668 G -> S (IN REF. 2).

SQ SEQUENCE 706 AA; 77612 MW; 2187A3C603EC7D39 CRC64;

Query Match 3.4%; Score 262; DB 1; Length 706;

Best Local Similarity 21.9%; Pred. No. 3.9e-05;

Matches 163; Conservative 106; Mismatches 285; Indels 190; Gaps 37;

QY 11 PSR-AETKRKRECPQOLPSPRR-----NTEKNREQENKYTEELAEILFANFNIDN 62

DB 29 PARGGKRRSGMDPDEDEGSKFSRENHSEIERRRNRKMQYITLSDM-----PTCSA 84

QY 63 FNFKPKCAIILKETVQKQIROIKECEKAAANIDEVQKSDVSTGGVIDKDALGPMLEA 122

DB 85 LARKDKLTILMAVSHHKSAR-----GTGN-----KSTDGAYKSFLEDEKLILIEA 134

QY 123 LDGFFPVNLE-GNVFVSENVTOYLRYNOBELMKNKSVYSLHVGDNTEPKNLPKSIV 181

DB 135 ADGFLEFVAAEGRVIYSDSVTPVLNQPSBMFESTLYEYHPPD-----VEKLEQGLCT 190

QY 182 NGGMSG-----EPPRRNS-----HFNCRMLVKKPLP-----DSEEE 213

DB 191 SENSMTGRILDLKTGTVKKEGQSSMRMCKGSRFICRMRCGAPLDLPLNRTTYMRK 250

QY 214 GHDN-----OEAHOKYETMOC-----FAVSQPKSITKEGEDLOSLICVARVPK 259

DB 251 RFRNGLGVKGEAOYAVNHCTGYIKAMPKPMGTIPEDDADVGQSKYLAIG-RLOYT 309

QY 260 ERPV-----LPSSSEFTTRQDLQKITSLLDTSTRAAMKPGW--EDLVRCIOKF-HAQ 310

DB 310 SSPVCMQDNKMSVPTFELSRHNSDGIITFVDP---RCISVIGYQPDLLGKDLLEFCRP- 365

QY 311 HEGEVSTAKKHHHYLR-QGLAFSQIYRFSLSQDTILVAADTKSLINSQTTNEQOLVI- 368

DB 366 ---EDQSHLRSPQOVVVKLGQVLSVMYRFRKKNEMWLITSSFTFQNPYSDELEYLIC 422

QY 369 ---SLHMLHROQ-----NVCVMNPDLTGQTMGKPLNPIS-----NSP 403

DB 423 TINTVKKQLOQOQDALELVHQRGLSSYDLSQYVPVRLPRAGVHACKSVKAKAITSQEDP 482

QY 404 AHQALCSGNPQODMTLSSNINPLNGPKQMG--MPMGRFCG--SGGNHVSQAQATTPQ 459

DB 483 RFAEFACIASASEKKMMSASAGAQTQIYSQGSFPFSGHSGKARSSVHVPGVY----- 537

QY 460 GSNVALKMNPSOSSPGNPNPOPTSMLSPRHRMSGVAGSPRIR---PSQFSPAGLSHS 515

DB 538 -----DIQSSSTGNGMSQISKROLNOSOVAMTG--SRPFPQOQIIPSOSSSTQO--S 585

QY 516 PVGVCSSTGNSHSYTNSLNLQALSLSEGHVSLGSSLASP--DLKMGNLGNSPVNMNPP 573

```

Db      586 PFGI-----GTSHTYPPADP-----SSYSPLSSTPATSPSCNAYSLANRP-- 626
Oy      574 LSKMSLSDKDCFGLY-GEPSEGTGQAESCHPGEOKEKENDPMLPAVSSERADQOSRL 632
Db      627 -----GFAESQSSGQFGGRSE-VMSQWQSQ-HHGQO-----SGEQHS 663
Oy      633 HDSKGQTKLLQLLTYSKSDMEPSP 656
Db      664 HQQPGTEVFQ-----DMLPMF 680

RESULT 24
ID      HIFA_MOUSE STANDARD; PRT: 822 AA.
HIFA_MOUSE
AC      061221; 061665; 061664; 008993; 008741;
DT      01-NOV-1997 (Rel. 35, Created)
DT      15-DEC-1998 (Rel. 37, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      Hypoxia-inducible factor 1 alpha (HIF-1 alpha) (ARNT interacting
GN      HIF1A.
OS      Mus musculus (Mouse).
OC      Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX      NCBI_Taxid=10090;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=C57BL/6; TISSUE=Hepatocytes;
RX      MEDLINE=96355491; PubMed=8702901;
RA      Li H., Ko H.P., Whitlock J.P. Jr.;
RT      Induction of phosphoglycerate kinase 1 gene expression by hypoxia.
RL      J. Biol. Chem. 271:21262-21267(1996).
RN      [2]
RP      SEQUENCE FROM N.A.
RC      STRAIN=129/SVJ;
RX      MEDLINE=98034461; PubMed=9368100;
RA      Luo G., Gu Y.-Z., Jain S., Chan W.K., Carr K.M., Hogenesch J.B.,
RA      Bradfield C.A.;
RT      Molecular characterization of the murine Hif-1 alpha locus.
RL      Gene Expr. 6:287-299(1997).
RN      [3]
RP      SEQUENCE FROM N.A.
RC      STRAIN=129/SV;
RX      MEDLINE=97354184; PubMed=9210478;
RA      Wenger R.H., Rolfs A., Kvietikova I., Spielmann P., Zimmermann D.R.,
RA      Gassmann M.;
RT      "The mouse gene for hypoxia-inducible factor-1alpha. Genomic
RT      organization, expression and characterization of an alternative first
RT      exon and 5' flanking sequence."
RL      Eur. J. Biochem. 246:155-165(1997).
RN      [4]
RP      SEQUENCE OF 13-822 FROM N.A.
RC      TISSUE=Hepatocytes;
RX      MEDLINE=96294028; PubMed=8660378;
RA      Wenger R.H., Rolfs A., Marti H.H., Guenet J.-L., Gassmann M.;
RT      "Nucleotide sequence, chromosomal assignment and mRNA expression of
RT      mouse hypoxia-inducible factor-1 alpha."
RL      Biochem. Biophys. Res. Commun. 223:54-59(1996).
RN      [5]
RP      SEQUENCE OF 22-85 FROM N.A.
RC      TISSUE=Hepatocytes;
RA      O'Rourke J.F.;
RL      Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
CC      -1- FUNCTION: INVOLVED IN THE INDUCTION OF OXYGEN REGULATED GENES.
CC      SPECIFICALLY RECOGNIZES AN 8 BP HYPOXIA RESPONSE ELEMENT (HRE).
CC      -1- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER
CC      BHLH PROTEIN. HETEROIMER OF AN ALPHA AND A BETA (ARNT) SUBUNITS.
CC      THE ALPHA SUBUNIT IS UNIQUE TO HIF-1 WHEREAS HIF-1 BETA (ARNT) CAN
CC      DIMERIZE WITH OTHER BHLH-PAS PROTEINS. INTERACTS WITH HSP90 (BY
CC      SIMILARITY).
CC      -1- SUBCELLULAR LOCATION: Nuclear (Potential).

```

```

CC      -1- TISSUE SPECIFICITY: UBQUITOUS.
CC      -1- DOMAIN: BOTH HYPOXIA RESPONSIVENESS AND TRANSACTIVATION CAPABILITY
CC      RESIDE WITHIN THE C-TERMINAL PART.
CC      -1- PTM: REQUIRES PHOSPHORYLATION FOR DNA-BINDING.
CC      -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
CC      TRANSCRIPTION FACTORS.
CC      -1- SIMILARITY: CONTAINS 2 PAS (PER-ARNT-SIM) DIMERIZATION DOMAINS.
CC      -1- SIMILARITY: CONTAINS 1 PAS-ASSOCIATED C-TERMINAL (PAC) DOMAIN.
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@sib-sib.ch).
CC
DR      EMBL: U59496; AAC52730.1; -.
DR      EMBL: AF003695; AAC53453.1; -.
DR      EMBL: Y09085; CAA70306.1; -.
DR      EMBL: Y13656; CAA70306.1; JOINED.
DR      EMBL: Y09085; CAA70305.1; -.
DR      EMBL: AF004155; AAC53461.1; -.
DR      EMBL: AF004141; AAC53461.1; JOINED.
DR      EMBL: AF004142; AAC53461.1; JOINED.
DR      EMBL: AF004143; AAC53461.1; JOINED.
DR      EMBL: AF004144; AAC53461.1; JOINED.
DR      EMBL: AF004145; AAC53461.1; JOINED.
DR      EMBL: AF004146; AAC53461.1; JOINED.
DR      EMBL: AF004147; AAC53461.1; JOINED.
DR      EMBL: AF004148; AAC53461.1; JOINED.
DR      EMBL: AF004149; AAC53461.1; JOINED.
DR      EMBL: AF004150; AAC53461.1; JOINED.
DR      EMBL: AF004151; AAC53461.1; JOINED.
DR      EMBL: AF004152; AAC53461.1; JOINED.
DR      EMBL: AF004153; AAC53461.1; JOINED.
DR      EMBL: AF004154; AAC53461.1; JOINED.
DR      EMBL: X95580; CAA64833.1; -.
DR      EMBL: X95002; CAA64458.1; -.
DR      MGI: 106918; H1fla.
DR      InterPro: IPR003015; HLH_Myc.
DR      InterPro: IPR001092; HLH_dlm.
DR      InterPro: IPR001321; HypoxIndf1A.
DR      InterPro: IPR001610; PAC.
DR      InterPro: IPR000014; PAS.
DR      Pfam: PF00785; PAC.1.
DR      Pfam: PF00989; PAS.2.
DR      PRINTS: PR01080; HYPOXIAF1A.
DR      SMART: SM00353; HLH.1.
DR      SMART: SM00086; PAC.1.
DR      SMART: SM00086; PAS.2.
DR      PROSITE: PS00038; HELIX_LOOP_HELIX; 1.
DR      PROSITE: PS50112; PAS.2.
KW      Repeat; DNA-binding; Nuclear protein; Transcription regulation;
KW      Activator; Phosphorylation.
FT      DNA_BIND 17 30
FT      DOMAIN 31 71 BASIC DOMAIN.
FT      DOMAIN 80 155 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).
FT      DOMAIN 228 298 PAS 1.
FT      DOMAIN 302 345 PAS 2.
FT      DOMAIN 31 31 PAC.
FT      CONFLICT 31 31 S -> T (IN REF. 1).
FT      CONFLICT 128 128 A -> T (IN REF. 2 AND 3).
FT      CONFLICT 351 351 I -> L (IN REF. 1).
FT      CONFLICT 511 511 E -> ERLQENVNTPNFSQ (IN REF. 2).
FT      CONFLICT 686 686 K -> N (IN REF. 3 AND 4).
FT      CONFLICT 785 785 E -> V (IN REF. 4).
SQ      SEQUENCE 822 AA; 91858 MW; E717ADCEAC9D795 CRC64;

```

Query Match 3.4%; Score 262; DB 1; Length 822;  
 Best Local Similarity 19.6%; Pred. No. 4.7e-05;  
 Matches 178; Conservative 132; Mismatches 329; Indels 268; Gaps 38;



```

0Y 123 LDGFEFVNVLE-GNVVPESEVNTQYLRNOBELMKNKSVSYLHNGDHTFEVKNLPSIV 161
Db 146 ADGELEFVVAAETGAVITYSDVSYLVLANOPSEMFSGSTLIECVHDD---VEKREDLCT 201
0Y 182 NGGSMWG-----EPPRRMS-----HTFCRMVLVPLP-----DSEEE 213
Db 202 SENSTJGTLDLKTGYVKEGQSSMRMCKMSRSLFCRMRCGNAPLDLPLRNITTMK 261
0Y 214 GHDN---QAHOKYETMOC-----FVSPKSIKEEGEDLOSCLICVARRPMK 259
Db 262 RFRMGLGPFVKEGEAOYAVVHTGYIKAMPAGMTIPEEDADVGGSYCYCAVIG-RLOVT 320
0Y 260 ERPV-----PPSESEFTTRDOLGKITSLDPTSMRAAMKPGW--EDLYBRCKOF--HAQ 310
Db 321 SSPVCMDMKSGMSVPEFLSRHNSGCIITFVP---KCLSVGYQPOLDLKDLIEFCHP- 376
0Y 311 HEGESVSYAKRHHHEVLR-QGLAFSQIYRESLSDGLTVAAOTKSKILRSQTNPEPOLVI- 368
Db 377 ---EDOSHRESFOOVVYKLGQVLSVMYREFTKRREMLILRTSFTFPQNSDEIEVYC 433
0Y 369 ---SLHMLHRRQ-----NVCVNNPDLQGTOMKPLNPISS-----NSP 403
Db 434 TMTNVKQLODQOQDAELEVHQRDGLSSYDLSQVRYPNPLPAGVHEAKSYEKADLFTSOERDP 493
0Y 404 AHOALCSGNPCODMTLSSNINEFPINGPEOMG--MPMGREG--SGGMNVSG---MOAT 456
Db 494 RFAEMFALIASSEKKMMSASASOSQIYISGSGFFPAGHSKAPASSVHVHPGVNDQSS 553
0Y 457 TPQGSNVYALKMNSPQSSPGMPCQOPPTSMLSPRHRMGBVGAAGSFRITPQSPSPAGSLASP 516
Db 554 SSTGQNI-----SQTSLROLNOCQ-----VAMTGSRRPFGQGPSKTOSSA 592
0Y 517 VGVCSSTGNSHSTYNNSSLNALQALSEHGVSIGSSLASP--DLKGMGLQNSPVMMNPP 574
Db 593 FGI-----GSSHPYPADP-----SSYPLSPAASSPGNAYPSLAKTP--- 632
0Y 575 SKMGLSDSKDCEFGYI-GEPSSEGTGOAESCHPEQOETKDPNPLPVPVSSBRADGQSRHL 633
Db 633 ---GRAESGGGQGGQFQGRPSE-VMSQMSQ-SHQGO-----SGEGHSH 670
0Y 634 DSKGOTKLQLDLITTKSDOMEPS 656
Db 671 QOPGQTEVFO-----DMLPMP 686

```

RESULT	26			
NPAA2_HUMAN				
ID	NPAA2_HUMAN	STANDARD:	PRT;	824 AA.
AC	Q997433, Q99629;			
DT	15-DEC-1998 (Rel. 37, Created)			
DT	15-DEC-1998 (Rel. 37, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Neuronal PAS domain protein 2 (Neuronal PAS2) (Member of PAS protein 4) (MOP4).			
DE	NPAS2.			
GN	NPAS2.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	(1)			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=97165088; Pubmed=9012850;			
RA	Shou Y.-D., Barnard M., Tian H., Li X., Ring H.Z., Francke U.,			
RT	"Molecular characterization of two mammalian bHLH-PAS domain proteins selectively expressed in the central nervous system.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 94:713-718(1997).			
RL	(2)			
RP	SEQUENCE OF 1-626 FROM N.A.			
RX	MEDLINE=97236817; Pubmed=9079689;			
RA	Hogensach J.B., Chan W.K., Jackiw V.H., Brown R.C., Gu Y.-Z.,			
RA	Pray-Grant M., Perdev G.H., Bradfield C.A.;			

RT	"Characterization of a subset of the basic-helix-loop-helix-PAS superfamily that interacts with components of the dioxin signaling pathway."
RL	J. Biol. Chem. 272:8581-8593(1997).
CC	-1 SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER BHLH PROTEIN. INTERACTS WITH HSP90.
CC	-1 SUBCELLULAR LOCATION: Nuclear (potential).
CC	-1 SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF TRANSCRIPTION FACTORS.
CC	-1 SIMILARITY: CONTAINS 2 PAS (PER-ARNT-SIM) DIMERIZATION DOMAINS.
CC	-1 SIMILARITY: CONTAINS 1 PAS-ASSOCIATED C-TERMINAL (PAC) DOMAIN.
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb.ch/announce/cor_send_an_email_to_license@isb-sib.ch">http://www.isb-sib.ch/announce/cor_send_an_email_to_license@isb-sib.ch</a> ).
CC	-----
DR	EMBL; U77970; AAB47250.1; -;
DR	EMBL; U51625; AAC51211.1; -;
DR	MIM; 603347; -;
DR	InterPro; IPR003015; HLH_MYC.
DR	InterPro; IPR001092; HLH_dlm.
DR	InterPro; IPR001610; PAC.
DR	InterPro; IPR000014; PAS.
DR	Pfam; PF00010; HLH; 1.
DR	Pfam; PF00785; PAC; 1.
DR	Pfam; PF00989; PAS; 2.
DR	SMART; SMO0353; HLH; 1.
DR	SMART; SMO0086; PAC; 1.
DR	SMART; SMO0091; PAS; 2.
DR	PROSITE; PS00038; HELIX_LOOP_HELIX; 1.
KW	PROSITE; PSS0112; PAS; 2.
RW	Repeat: DNA-binding; Nuclear protein; Transcription regulation.
FT	DNA_BIND 10 22 BASIC DOMAIN.
FT	DOMAIN 23 60 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).
FT	DOMAIN 82 152 PAS 1.
FT	DOMAIN 237 307 PAS 2.
FT	DOMAIN 311 354 PAC.
FT	CONFLICT 51 K -> E (IN REF. 2).
FT	CONFLICT 164 S -> G (IN REF. 2).
FT	CONFLICT 308 T -> K (IN REF. 2).
FT	CONFLICT 471 L -> S (IN REF. 2).
SQ	824 AA; 91/759 MW; 249AAC46b7B328a5 CRC64;

	Query Match	3.48:	Score 259:	DB 1:	Length 824:	
	Best Local Similarity	19.48:	Pred. No. 6.4e-05:			
	Matches 215:	Conservative 154:	Mismatches 399:	Indels 338:	Gaps	46:
OY	4	MGENTSDPSRAETRRKRCPPDLOGSPKRNETRRENOENKYIEELAEIAPANFNIDINF	63			
Dd	1	MDEEDKAARAKASRK-----SEKKRRDPFWYLIKLSMLGNFR-----	41			
OY	64	NFKPDCAILKETYQRIQIRKEEKAANIDEVOKSDVSSTGOGVIDKDALGPMTLEAL	123			
Dd	42	--KMDTKTVEKVIGFLG--KHNEVSAQTETCIDIQDDMKPS---FLNSEPTQGLMEAL	93			
OY	124	DGEFFVVNLGEGNVVFVSENATQYLYRYNOBELMNSVSYLHVGDTTFVKALLPKRSYNG	183			
Dd	94	DGFIAVTITDSISIIIVSSITPRLIGHLPSPDWDOMLTLNPLQEHESEVYKILLSHMLVTD	153			
OY	184	GSMGSEPPRRNSHT---FNCRMVLKPYLPDSEEBGD-----NQEAHQYETMQC----F	230			
Dd	154	--SPSPYYLKSDSDFCYCHLLRGSLNPKEFPPTYEYIKFGNFESYNVPSPSCNGFDN	210			
OY	231	AVSQOSKKEGEDJOSGLICVARVP---KKERVLPSS--ESPTRDLODKITSLPTS	286			
Dd	211	TLSRCRV--PLGKEY--CFIATVRLATQFTLKEMCIYDEPILEBFTSRSLSEKFLFLD--	265			
OY	287	TMRAMKRGW---EDLVRCIOKFPAQHGEGSVYAKRHNHFEVLRNGLAFSAOIRFSLSD	343			

```

Db      266 -HRAPIIGLYFEVLGTSGYDYHI-----DDLELLAROHQMLMGFGTGKSCCYRFLTRKG 320
Oy      344 GTLVAAQTKSKIRISQOTINEPOLVISTLMLHREQVNCVNPDLTQOTMCKPLNPISNSP 403
Db      321 QQWMLQHYHYTYTHQWMSKPEFIVCTHVSYSYADV-----E 359
Oy      404 AHQALCSNPGQDMTSSININFPINPKPEQMMPMGREG-----GSGGMHVSQMQATPQ 459
Db      360 RQDELALDPPEALHSSALK-----DKGSSLEPQHFNALDVGASGLN-----403
Oy      460 GSNVALKNMSPSSPGMNPQPTSMLSPRHMSR--GVAGSPRIIPPSQSPAGSLHSPV 517
Db      404 -----TSHSPSASSRSHKSSHTAMSEPTPTPKLMAEASPALPRSATLPQ---ELPV 454
Oy      518 GVCSTGSHSTNSLNAALALSGHGVSLGSSLASP---DLKNGNLQNSVYNNMPPL 574
Db      455 -----PGLSQ--AATMPAPLPSPSLCDLTQOLLPOTVLQSTPAPM 492
Oy      575 SKMGSLDKDFGLGEPSEGTGQAESSCHPGEQKETNDPPLPAVSESRAD--GOSR 631
Db      493 AQFSNQ-----FSMF-----QTIKQLEORTRLQANIMQDEE 526
Oy      632 LHDSKGQTKLLQLLTTKSDQNEPPLASLSLDTNKSSTGSLPGSGSTHGTSLKEKHKLH 691
Db      527 LHKIQEQLCLVQDSNVQMFLOQPA-VSLFSSTQRPFA-----Q 564
Oy      692 RLDDSSSPVDLAKLTAEATGKDLQSESSSTAPGEVITIKOEPV-----SPKKENALLRY 747
Db      555 QQLQQRSAAVQPOLGA---GPQLPGQISSAQVTISQHLRESSVISTOGPKPMRSSQLMQ 621
Oy      748 -----LDDKDKTDIGLPEITPKLERLDSKTDPASTYKLIAMKTEKEKSFPGDOP 799
Db      622 SSGRGSLLVEPFSSATAALP---PSL-----NLTPPASTS-----QD 656
Oy      800 GSEIDNLEIILDDQLNSQLPQLFPDTPRGAPAGAVDKQAIIINDLMQLTAENSPTVPVGAQ 859
Db      657 ASQCPSPDFSHDRLRL--LSQPIQPMMP--GSCDAR-----QPEVSRGTGQ 702
Oy      860 KTAIRISQSTNNRRPGQLGRILRNQNLPLDTTILSPGAGFPPIRNSPVSVPQGM 919
Db      703 VKYAO-SQTVFQNP-----DAHPANSSAPMPVL-----730
Oy      920 MGNQMGIONGNLGNSTGMIGNSASRPTMPSGEWAPQSSAVRYCAATTSAMNRPVGG 979
Db      731 -----LMQAVLHPSFASQ-----745
Oy      980 MIRNPASIPMRPSSQPGQROTILQSOVANNIGPSELEMMNGGPOYSQ-----QAPPNQ 1032
Db      746 ---PSPHQPAQARQOPQPHY-LQVQAPTLSHSEQDILLSTYSQQFGTLGYPQPPAQ 800
Oy      1033 TAPWESILPIDQASFASQ---RQP 1055
Db      801 ---PQPLRPFRVSSLSSESGLOQP 822

```

```

RT      "Involvement of a homolog of Drosophila trithorax by 11q23
RT      Chromosomal translocations in acute leukemias.";
RT      Cell 71:691-700(1992).
RN      [2]
RP      SEQUENCE FROM N.A.
RX      MEDLINE-96290553; PubMed-8703835;
RA      Nilsson I., Loechner K., Slegler G., Grell J., Beck J.D., Fey G.H.,
RA      Marschalek R.;
RT      "Exon/intron structure of the human ALL-1 (MLL) gene involved in
RT      translocations to chromosomal region 11q23 and acute leukaemias.";
RT      Br. J. Haematol. 93:966-972(1996).
RN      [3]
RP      SEQUENCE OF 1-1909 FROM N.A.
RX      MEDLINE-93390935; PubMed-8378076;
RA      Yamamoto K., Seto M., Komatsu H., Iida S., Akao Y., Kojima S.,
RA      Kodera Y., Nakazawa S., Ariyoshi Y., Takahashi T., Ueda R.;
RT      "Two distinct portions of LTG19/BNL at 19p13 are involved in t(11;19)
RT      leukemia.";
RT      Oncogene 8:2617-2625(1993).
RN      [4]
RP      SEQUENCE OF 1317-2328 FROM N.A.
RC      TISUE-Brain;
RX      MEDLINE-93265134; PubMed-1303259;
RA      Djebail M., Salleri L., Parry P., Bower M., Young B.D., Evans G.A.;
RT      "A trithorax-like gene is interrupted by chromosome 11q23
RT      translocations in acute leukaemias.";
RT      Nat. Genet. 2:113-118(1992).
RN      [5]
RP      SEQUENCE OF 1251-1538 FROM N.A.
RX      MEDLINE-94215165; PubMed-8162575;
RA      Gu Y., Alder H., Nakamura T., Schlichman S.A., Prasad R., Canaanl O.,
RA      Saito H., Croce C.M., Canaanl E.;
RT      "Sequence analysis of the breakpoint cluster region in the ALL-1 gene
RT      involved in acute leukemia.";
RT      Cancer Res. 54:2326-2330(1994).
RN      [6]
RP      SEQUENCE OF 1251-1654 FROM N.A. (ISOFORM 14P-18B).
RX      MEDLINE-9532025; PubMed-7598802;
RA      Mbongkolli D., Burnett R., McCabe N., Thirman M., Gill H., Yu H.,
RA      Rowley J.D., Diaz M.O.;
RT      "The human MLL gene: nucleotide sequence, homology to the Drosophila
RT      trt zinc-finger domain, and alternative splicing.";
RT      DNA Cell Biol. 14:475-483(1995).
RN      [7]
RP      SEQUENCE OF 1212-1603 FROM N.A.
RX      MEDLINE-95315013; PubMed-7794749;
RA      Marschalek R., Grell J., Loechner K., Nilsson I., Slegler G.,
RA      Zweckbronner I., Beck J.D., Fey G.H.;
RT      "Molecular analysis of the chromosomal breakpoint and fusion
RT      transcripts in the acute Lymphoblastic SEM cell line with chromosomal
RT      translocation t(4;11).";
RT      Br. J. Haematol. 90:308-320(1995).
RN      [8]
RP      SEQUENCE OF 1421-1540 FROM N.A.
RX      MEDLINE-94020842; PubMed-8414518;
RA      Forster A., Rablitts T.H.;
RT      "A method for identifying genes within yeast artificial chromosomes:
RT      application to isolation of MLL fusion cDNAs from acute leukaemia
RT      translocations.";
RT      Oncogene 8:3157-3160(1993).
RN      [9]
RP      CHROMOSOMAL TRANSLOCATION WITH GAS7.
RX      MEDLINE-20183971; PubMed-10706619;
RA      Megonigal M.D., Cheung N.-K.V., Rappaport E.F., Nowell P.C.,
RA      Wilson R.B., Jones D.H., Addya K., Leonard D.G.B., Kushner B.H.,
RA      Williams T.M., Lange B.J., Felix C.A.;
RT      "Detection of leukemia-associated MLL-GAS7 translocation early during
RT      chemotherapy with DNA topoisomerase II inhibitors.";
RT      Proc. Natl. Acad. Sci. U.S.A. 97:2814-2819(2000).
RN      [10]
RP      CHROMOSOMAL TRANSLOCATION WITH AF3P21.
RX      MEDLINE-20115194; PubMed-10648423;
RA      Sano K., Hayakawa A., Piao J.-H., Kosaka Y., Nakamura H.;

```

RT Novel SH3 protein encoded by the AF3p21 gene is fused to the mixed  
 RT lineage leukemia protein in a therapy-related leukemia with  
 RT t(3;11)(p21;q23).  
 RL Blood 95:1066-1068(2000).  
 CC -1- FUNCTION: POSSIBLY ACTS AS A TRANSCRIPTIONAL REGULATORY FACTOR.  
 CC -1- SUBCELLULAR LOCATION: Nuclear.  
 CC -1- TISSUE SPECIFICITY: HEART, LUNG, BRAIN AND T AND B LYMPHOCYTES.  
 CC -1- DISEASE: INVOLVED IN ACUTE LEUKEMIAS BY CHROMOSOMAL TRANSLOCATIONS  
 CC T(11;19)(Q23;P13.3) THAT INVOLVES MLL AND MLT1/ENL;  
 CC T(4;11)(Q21;Q23) THAT INVOLVES MLL AND MLT2/AF4; T(9;11)(P22;Q23)  
 CC THAT INVOLVES MLL AND MLT2/AF9; T(6;11)(Q27;Q23) THAT INVOLVES  
 CC MLL AND MLT4/AF6; T(11;17)(Q23;Q21) THAT INVOLVES MLL AND  
 CC MLT6/AF17; T(X;11)(Q13;Q23) THAT INVOLVES MLL AND MLT7/AFX1;  
 CC T(10;11)(P12;Q23) THAT INVOLVES MLL AND MLT10/AF10;  
 CC T(1;11)(Q21;Q23) THAT INVOLVES MLL AND AF10; T(11;19)(Q23;P13.3)  
 CC that involves MLL and ELL; t(11;19)(q23;p23) that involves MLL  
 CC and GAS7; and t(3;11)(p21;q23) that involves MLL and AF3p21.  
 CC -1- SIMILARITY: BELONGS TO THE TRANSCRIPTION FACTOR TRITHORAX FAMILY.  
 CC -1- SIMILARITY: CONTAINS 1 BROMODOMAIN.  
 CC -1- SIMILARITY: CONTAINS 3 PHD-TYPE ZINC FINGERS.  
 CC -1- SIMILARITY: CONTAINS 1 CXXC-TYPE ZINC FINGER.  
 CC -----  
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: L04284; AAAS8669.1; -;  
 DR EMBL: 269744; CAA93625.1; -;  
 DR EMBL: 269745; CAA93625.1; JOINED.  
 DR EMBL: 269746; CAA93625.1; JOINED.  
 DR EMBL: 269747; CAA93625.1; JOINED.  
 DR EMBL: 269748; CAA93625.1; JOINED.  
 DR EMBL: 269749; CAA93625.1; JOINED.  
 DR EMBL: 269750; CAA93625.1; JOINED.  
 DR EMBL: 269751; CAA93625.1; JOINED.  
 DR EMBL: 269752; CAA93625.1; JOINED.  
 DR EMBL: 269753; CAA93625.1; JOINED.  
 DR EMBL: 269754; CAA93625.1; JOINED.  
 DR EMBL: 269755; CAA93625.1; JOINED.  
 DR EMBL: 269756; CAA93625.1; JOINED.  
 DR EMBL: 269757; CAA93625.1; JOINED.  
 DR EMBL: 269758; CAA93625.1; JOINED.  
 DR EMBL: 269759; CAA93625.1; JOINED.  
 DR EMBL: 269760; CAA93625.1; JOINED.  
 DR EMBL: 269761; CAA93625.1; JOINED.  
 DR EMBL: 269762; CAA93625.1; JOINED.  
 DR EMBL: 269763; CAA93625.1; JOINED.  
 DR EMBL: 269764; CAA93625.1; JOINED.  
 DR EMBL: 269765; CAA93625.1; JOINED.  
 DR EMBL: 269766; CAA93625.1; JOINED.  
 DR EMBL: 269767; CAA93625.1; JOINED.  
 DR EMBL: 269768; CAA93625.1; JOINED.  
 DR EMBL: 269769; CAA93625.1; JOINED.  
 DR EMBL: 269770; CAA93625.1; JOINED.  
 DR EMBL: 269772; CAA93625.1; JOINED.  
 DR EMBL: 269773; CAA93625.1; JOINED.  
 DR EMBL: 269774; CAA93625.1; JOINED.  
 DR EMBL: 269775; CAA93625.1; JOINED.  
 DR EMBL: 269776; CAA93625.1; JOINED.  
 DR EMBL: 269777; CAA93625.1; JOINED.  
 DR EMBL: 269778; CAA93625.1; JOINED.  
 DR EMBL: 269779; CAA93625.1; JOINED.  
 DR EMBL: 269780; CAA93625.1; JOINED.  
 DR EMBL: D14540; BAA03407.1; -;  
 DR EMBL: L01986; AAA92511.1; -;  
 DR EMBL: U04737; AAA18644.1; -;  
 DR EMBL: S78570; AAB34770.1; -;  
 DR EMBL: X83604; CAA58584.1; -;

DR EMBL: S66432; AAB28545.1; -;  
 DR EMBL: AF231998; AAG26332.2; ALT\_TERM.  
 DR TRANSFAC: T02337; -;  
 DR MIM: 159555; -;  
 DR InterPro: IPR001487; Bromodomain.  
 DR InterPro: IPR003889; Fyrich.C.  
 DR InterPro: IPR003888; Fyrich.N.  
 DR InterPro: IPR001965; PHD.  
 DR InterPro: IPR003616; PostSET.  
 DR InterPro: IPR001214; SET.  
 DR InterPro: IPR002857; znf-CXXC.  
 DR Pfam: PF00628; PHD; 3.  
 DR Pfam: PF00856; SET; 1.  
 DR Pfam: PF02008; zf-CXXC; 1.  
 DR SMART: SM00297; BROMO; 1.  
 DR SMART: SM00542; FYR1; 1.  
 DR SMART: SM00541; FYR2; 1.  
 DR SMART: SM00540; PHD; 4.  
 DR SMART: SM00508; PostSET; 1.  
 DR SMART: SM00317; SET; 1.  
 DR PROSITE: PS50014; BROMODOMAIN\_2; 1.  
 DR PROSITE: PS50280; SET; 1.  
 KW Proto-oncogene; Chromosomal translocation; DNA-binding; Bromodomain;  
 KW Nuclear protein; Zinc-finger; Metal-binding; Transcription regulation;  
 KW Alternative splicing.  
 FT DOMAIN 17 102  
 FT DNA\_BIND 169 180  
 FT DNA\_BIND 217 227  
 FT DNA\_BIND 301 309  
 FT ZN\_FING 1147 1194  
 FT ZN\_FING 1431 1482  
 FT ZN\_FING 1484 1533  
 FT ZN\_FING 1566 1627  
 FT DOMAIN 1703 1748  
 FT DOMAIN 3840 3969  
 FT DOMAIN 137 143  
 FT DOMAIN 561 564  
 FT DOMAIN 568 571  
 FT SITE 1362 1363  
 Query Match 3.4%; Score 258.5; DB 1; Length 3969;  
 Best local Similarity 18.6%; Pred. No. 0.0005;  
 Matches 350; Conservative 243; Mismatches 611; Indels 673; Gaps 91;  
 QY 10 DPSRAETRRKRCPCDQGPSKRNTERKNEQENKYLEELAFANFDIDNFKPK 69  
 DB 1673 NPETERSIPSRSSPE--GPDPVLTVEVSKDDQPL-----DLSEVKKRMQ 1717  
 QY 70 ---CALIK---ETVKQIQKQKQKAAANID---EVQKSDVSSGCGVIDKDALGPM 119  
 DB 1718 GNYTSVLEFSDDIVKIIQ-----AAINS DGQPEIRKAN-----SM 1753  
 QY 120 LEALDFFEFVNEGVVSENVTOYLRYNOELNNKSVSYSLHVGDTPEYKN-LLPK 178  
 DB 1754 VKS-----FFIRQMERVPFPMVSKSRFMBENK-----VSSNSGMLPNAVLPP 1796  
 QY 179 SI-VNGSGMSGEPDRNSHTFNCRLVPLPDSSEEGHNDQEAHQYETMQCFVSPKRS 237  
 DB 1797 SLDHNYAOW--QERENSHTEQPLMKKIIPAKPKPGGPDSPPLPHPTPLSLDRS 1854  
 QY 238 IKEEG-----EDIQSLICV-----ARRVPAKERP-----VLPSSEFTTR 273  
 DB 1855 REDSPELNPPGIEDNRQOCALCTYGDDSDANDAGRLLYIGQENMTVNCALMSAEVFE 1911  
 QY 274 QDQOGKITSLDSTMTBA-----AMKPGWE--DLVRRCIQKHQAHE-----GESV 316  
 DB 1912 EDDDGSLKNVHMAVYNGKQRCFECPKPGATVCCCLTCTSNYHPCSRANKCVFLDDXK 1971  
 QY 317 SYAKRRH-----HEVLRQ-----GLAFSQIYRF----- 339  
 DB 1972 VYQGRHDLIKGEVPENGEVEFRVYFVEDEGTSLRKRLFLNGLEPENIHMMIGSMITDCL 2031  
 QY 340 ----SLSDGTLVAQTK-----SKLIRSQTTEPQVLYISLHMLHREQNVCMNPDLT 387



```

Db 2033 GILNDLSD-----CEDKLEPIGYOCSRYWSTTDARKRCVYCKIV--ECRPVVEPDI- 2083
OY 388 GQTMKPLNPISSNSPAHALCNSGPNQGMOTLSSNINFPINPKQGMKPMKRFQSSGM 447
Db 2084 NSTVHEDEKRTIAHSPSTSESSKESONTA-----EISPSDDRPHSOTSGCY 2136
OY 448 NHVSGM-QATTPQSGNYALKMNSPQSSPGMNP-----GQPTSLMSPRHR-----MS 494
Db 2137 HVIKSVPRIRTPS-----YSPQSRSGCRPLPSAGSPT---PTHETIVYVGPDLSS 2185
OY 495 GV--AGSPRIIPSPSPACS---LHSPVGCSTGNSHSYSTNS-----533
Db 2186 GLRSIGSRHSTSLSPQSKLRIMSPM---RTGNTYSRNNVSVSTGTATDLESSAK 2241
OY 534 -----LINAQALSEGCVGLSSLASPDLMKMLQNSPVMNMPRLSKNGS---LDS 582
Db 2242 VVDHVLGPNLNSSTL--GQNTSTSSMLQRTVVTGVN--KNSHLDGSSSSSEMKOSSASDLVS 2298
OY 583 KDCFGLYGE-----PSEGTGQAESSCHPEQKETNDPMLPPAVSERADGOSRLHDSKG 637
Db 2299 KSS--SLKGEKTKVLSSKSESSAHNVAYPG-----IPLKAP-----2333
OY 638 QTKLIQLLTYSQDMEPPLA-----SSLDTNKGSTGSLPGSGSTHGTSLKEKRIIHR 692
Db 2334 -----QVHNTTSRELNVSIGSFAPSSSVSFSSKE-----ALSFPHLHLRG 2374
OY 693 LLQDSSSPVDLAKLFAEATGKDLSSSESTAGSEVTIKQEPVSPKKKNALLRYLLMD 752
Db 2375 QNRNDQDHTD-----STQSANSPDEDETEVTKLISGNSNRSI-----2414
OY 753 DTKDIGLEPITKLERLDSKTPASTNFKLIAMKTEKEKS-----PEPQDQSELDNLE- 807
Db 2415 -NEHMG-----SSRRROKKGKSKCKETREKIKSSSFLERQVTTGEBENLKP 2462
OY 808 EILDDLQNSQLPOLFPD-----TRGPAP-----AGSYD-----835
Db 2463 BEMDEV-----LTPEYWGQRPCCNNVSSDKICDKGLSMGPVKAPPMQVESSAKELQAP 2515
OY 836 -KQALINDLMQLTAE-----SPVTVGAKQKTLRLISQSTFNPRGQGLRLP 883
Db 2516 KRRYKVTLLTKMENESOSKNALKESSPASPLQIESTSPTEPISASENPDGQVAAQSP 2575
OY 884 N-----QNLPL-DITLQSPGAGP-----FP-----PIR 906
Db 2576 NNTSQDOSQSNYQNLPLPQDRMLMLPDGPKPQEDGSKRRYRRSARASNMFPGLTPLY 2635
OY 907 NSSPY-----SVLPQGMGNQGMIGNQNLGNS-----STGMIG 941
Db 2636 GVRSYGEEDIPYSSSTGKRGKRSAGQVDGADLSTDEDDLYYNTFTVVISGGE 2695
OY 942 NSASR-----PRMPSGEKAPQSSAVAVTCAATTSAMNR-PVQCGMIRNPASTP 989
Db 2696 RLASHNLFREBQCDLPKISQLDGVDDGTESDTSVATTRKSSQILPKRNGK-ENGLENLK 2754
OY 990 MRPSQPOROTLO-----SOVNIGPSELENMGQPOYSQ-----1025
Db 2755 IDRPDAGEKEHVTVSSVGHKNEPKMDCNCHSVRKVQGOQDLEAQLSSLESSRKYHST 2814
OY 1026 -----QOAPPNQTAWPMSILPDIQASFAQSN---ROPGSSPDDLCPH 1067
Db 2815 PSDKMLDPTYNTELKSSDNNNSDDCGNILLPSDIMDFLKNTPSMQALGESP-----2867
OY 1068 PAESPSPDEGALLDQLYALRNFDLE-EIDRALGIPELVSOQAVDPEQSSQDSNIML 1126
Db 2868 ---ESSSSB-----LNLGEGLDISNREKDMGLFEVFSQQLPTTEBVDSSVSSISA 2917
OY 1127 BO-----KAPVPOQYASQAMQAGVSPQODNFHMGOR-----P 1163
Db 2918 EQQFLEPLEPLPSDLVLTTRSPVSSQNSRLAVISDS-----GKRVYITTEK 2965
OY 1164 SVATLRMP-----RPLRPTGLVONOPNQLRLQRLAQOONROPLMNOI--SNVSNVL 1218

```

```

Db 2966 SVASSESDPALLSPGVDPPTPEGHMPDHF--IOGHMADHISPPCGSVQGHGNNDDL 3022
OY 1219 TLRPGVP--TQADIPNQMLOARORELNO-----HLR---OROMHQ 1255
Db 3023 TRNNSSTPGLOVPEVSPVPIQONOKIVPNSDPSQPOISANAQVOTPPHKLPAKTELIVYN 3082
OY 1256 QOVQOQ-----RTL-----MARGOGLNMT-----SMVASGPMATMS---NPRIPA 1294
Db 3083 QMMQPLVYLQTLPNQVYQKIQTLSSVSSPVSMENTVSLVGMGGGLTLITGLNPLSPRS 3142
OY 1295 N-----AQOPFPNN-----YGISQOPDPEF-----1315
Db 3143 QSLFSPASKGLLPMSHHQHLHSPPAATOSFPFNISNPPGILIGVQPPDQLLVSESS 3202
OY 1316 -----TGATTPQSPMLSPRMAHTQSPMMQO---SQANPAYQAPSDI-----1353
Db 3203 QRTDLSITVATPSSGGLKRPISRLQTRKKKLAPSPSPSNIPADSVSMMLINFTPSOL 3262
OY 1354 -NGMAQNGMNGSMFESQSPPHGOQANT-S-MYSNNMNIIVSMATNTGSMSSMNQTCOI 1411
Db 3263 PNHPLSLDLGSLNLTSSHRTVPNIRKSKSINVEPAPL---LPQSVGCTAFAAGTSTI 3319
OY 1412 SM--TSVTSVSTGLSS 1426
Db 3320 SODTSHLTSGSVGLAS 3336

RESULT 28
MOZ_HUMAN STANDARD; PRT; 2004 AA.
ID 092794;
AC 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Monocytic leukemia zinc finger protein (zinc finger protein 220).
GN ZNF220 OR MOZ.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96376968; PubMed=8782817;
RA Borrow J., Stanton V.P., Jr., Andresen J.M., Becher R., Behn F.G.,
RA Chaganti R.S.K., Civin C.I., Distche C., Dube I., Fritschau A.M.,
RA Housman D., Mittleman F., Volinia S., Watson A.E., Housman D.E.,
RT "The translocation t(8;16)(p11;p13) of acute myeloid leukemia fuses
RT a putative acetyltransferase to the CREB-binding protein."
RT Nat. Genet. 14:33-41(1996).
CC -!- FUNCTION: MAY REPRESENT A CHROMATIN-ASSOCIATED ACETYLTRANSFERASE.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- DISEASE: PARTICIPATES IN A T(8;16)(p11;p13) CHROMOSOMAL
CC TRANSLOCATION THAT PRODUCES A MOZ-CBP CHIMERA OBSERVED IN THE
CC M4/M5 SUBTYPE OF ACUTE MYELOID LEUKEMIA.
CC -!- SIMILARITY: CONTAINS 2 PHD-TYPE ZINC FINGERS.
CC -!- SIMILARITY: BELONGS TO THE MYST (SAS/MOZ) FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U47742; AAC50662.1; -.
CC HSSP: 060631; 4GBQ.
CC
CC MIM: 601408; -.
CC
CC InterPro: IPR001386; Linker_histone.
CC InterPro: IPR002717; MOZ_SAS.
CC InterPro: IPR001965; PHD.
CC Pfam: PF01853; MOZ_SAS; 1.
CC Pfam: PF00628; PHD; 2.

```

DR SMART; SM00526; H15; 1.  
 DR SMART; SM00249; PHD; 2.  
 KW Proto-oncogene; Chromosomal translocation; Zinc-finger; Repeat;  
 KM Nuclear protein.  
 FT ZN\_FING 206 256 PHD-TYPE 1.  
 FT ZN\_FING 259 313 PHD-TYPE 2.  
 FT DOMAIN 371 379 POLY-SER.  
 FT ZN\_FING 538 560 C2HC-TYPE.  
 FT DOMAIN 788 801 POLY-GLU.  
 FT DOMAIN 989 995 POLY-GLU.  
 FT DOMAIN 1019 1026 POLY-ARG.  
 FT DOMAIN 1069 1078 POLY-GLU.  
 FT DOMAIN 1147 1150 POLY-LYS.  
 FT DOMAIN 1221 1242 GLU-RICH.  
 FT DOMAIN 1267 1302 GLU-RICH.  
 FT DOMAIN 1411 1414 POLY-GLU.  
 FT DOMAIN 1593 1597 POLY-SER.  
 FT DOMAIN 1643 1704 GLN/PRO-RICH.  
 FT SITE 1897 1977 MET-RICH.  
 FT SITE 1546 1547 BREAKPOINT FOR TRANSLATION TO FORM MOZ-CBP.  
 SO SEQUENCE 2004 AA; 225054 MW; 9FFBAC3792854BA CRC64;

Query Match 3.4%; Score 258; DB 1; Length 2004;

Best Local Similarity 19.4%; Pred. No. 0.00022;

Matches 264; Conservative 168; Mismatches 574; Indels 354; Gaps 61;

QY 169 TEFVKNLLPSIV-----NGSWSGEPFRNRSHTF---NCRMLVRLDSEEGDNDQ 218  
 DB 844 TRISKVLPDLSLPLANSQPSRRGRW-GRKNKTKQERFGDKSLLETSSAPQEQG-- 900  
 QY 219 EAHQVETMOCFAVQPSKIEGEDLOSLICVARRVPKEREVLPSSESFTTRDLOG 278  
 DB 901 ECGKRSATQ-----EQTSEEDLVASEEQSDGKRDLPKRRLSEGEVWRG 949  
 QY 279 KITSLDSTYRAAMKPCWEDLVRRICQFHAQHGESVSTAKRHHEVLAQGLAFSQTIR 338  
 DB 950 QLRK-SPEAKKCRLTGEGSERLPRYSGDAVALRGFSESSEEEEPSPR----- 998  
 QY 339 FSLSDGLVAAGPQSK---LIRSQTNPOLVLSLHMLHREONCVANPDLTGTGMPKL 395  
 DB 999 -SSPPLTLPTLKRKKRPFLLRRRRVAKK-----HNSSVYTETISETTEVLEDPF 1049  
 QY 396 NPISNSPAHQALCSGNPGQDMTLSSNINPINGPKDMG---MPGRRFGSGGMHVS 451  
 DB 1050 EDSSEPRM-----PRLEPT-----FEIDEEEDENELFPREYF-----RRLS 1089  
 QY 452 GMAQTTQGSNYALKMSPQSSFGMNPQPTSLSP-----RHMSGCVAGSPRIIPSOQ 507  
 DB 1090 SQDYLRCQSS--SKRKSDEEEDSDADDTPLKPVSLLRK--DVKNSPLEPDTST 1144  
 QY 508 S-----PAGSLSPVGVCSSTGNSHTYNSLALQALSEGVAISGLASLPD--- 557  
 DB 1145 PLKKKKWPKGKSRKPLHMKKRPKRGFKLS-----REIMPVTOACVIEPIYSIP 1196  
 QY 558 KMG---NLQNSPVNMNP---PLSKMSLSDKCGFLYGPSEGTGQ----- 599  
 DB 1197 KAGKRPKIQSESETEVEPKEDMLPE---ERKEEEMQAEAEAEAEEDDASSEVPAA 1252  
 QY 600 --AESSCHPGEQKTNPNPLPRAVSSFRADQGRHLHSDKQTKLLQTLTKSDQMESP 657  
 DB 1253 SPADSSNSP--ETETKPEVEEEEEKPRVSEEQROSEEQO-----ELEPEPEEE 1301  
 QY 658 ASSLSDNKSDSTGSLPSSGSHGSLKKEHKILHRLDQSSPYDLAKLTAATGKLSQ 717  
 DB 1302 EDAAEATQNDHDHADDDEGHLESTKK-----ELEQPIREYVKE 1343  
 QY 718 ESSSTAGSEVTTIKQEPVSPKKENALLRLLDKDQTKDGLPETPK-----LERL-- 769  
 DB 1344 E-----PGVQESFLDANMQKSRK-----IKDKEET-ELDSEEGQSHDTSVYSEQMAG 1391  
 QY 770 ---DSKTDPAASNKLKLMKTEKEMSFEPGDQGESLQNLDEI-----LDDLQNSQLPOL 821

DB 1392 SEDDHEDSHTKBELLELKEEEE-----IHSLELD-LETQVAOVSQLEQSSSHEGA 1442  
 QY 822 FPD-----TRPGAPASVQKQATINDLMQULTAENSVPTVGAQKTALRISQTFNN 872  
 DB 1443 YQDCEETLACQTLQSTQADEDPQMSVWDCHASSENSTISSV-----Q 1487  
 QY 873 PRGQGLRLLPQNLPL--DITLQSP-TGAGFPPIRNSPPSVIIPQGMGNQMGIN 928  
 DB 1488 SHPSQSVRSVSPNVPALESQGTQISPEQGSLSAPSNQNETSPMDVPSVSHSQOYVD 1547  
 QY 929 QG-----NLGNSST-----GMIGSASRPMPSEBAPQSSAVVYTCATIT 969  
 DB 1548 SGFSDLSISTETENENPSSYSTMGSSICGNSSSQSCSYGLSSSLTQSSCVVYQ 1607  
 QY 970 SAMNRPVQGMIRNPASISMPRSSQPGQRQT---OSQVMNIGPSELENNMGPGYISQ 1026  
 DB 1608 QMAS-----MGSSCMQOQSSVQPANCSIKSPQSCVERPPSNQOQPPPPPPQPP 1659  
 QY 1027 QAPPNQTAPEBESILPIDQAFASQNRQPPGSSPDDLCPHAPAESPSDEGALLDQLYLA 1086  
 DB 1660 QPPPPQPPAPQPPPPQPPQ---PQQQPPQPPQPPPP---PPPPQOQPP-----LSQCSMN 1708  
 QY 1087 LRNFQGLEETIDRALGIPELVQSQAANDPEQFSSQDSNIMLEQKAPVPPOQYASQAQMAQ 1146  
 DB 1709 ---NSFTPAPMIMEIPE-----SGSTGMSISTYERIP-----GDFGAG 1742  
 QY 1147 SYS-----PMODPNFHTMGQRP---SYATLRMQPPRGLPTGLYONQNP 1187  
 DB 1743 SYQPSATFSLAKLOQTLNTIMDPHAMPIHSAPVYSYAT-----SVSLNTGLAQAPS 1797  
 QY 1188 QLRLOLQRLQAOQNRQPLNQISNVSNVLT-----RCQVPTQADINQMLAQRO 1239  
 DB 1798 H---PLAGTPQAQATMPTPPLNLASTMTNLSPILQCMASATNIGIPHTQLQCOMPKGH 1854  
 QY 1240 RELINQH--LRQOMHQOQVOQRTLMRGQGLMTPSMYAPSGMRTMSPNPTIPQANAQ 1297  
 DB 1855 ISIRKSAAPLSAANHQO-----LYGR---SPSAVMAQAPRALAVQR----- 1895  
 QY 1298 QPFPPNYVIGISQPDGFTGATTPQSPILMSPRMAHNPQPMQOS-QANPAY-----QAPS 1351  
 DB 1896 -----GMMGMVNLMTPTAYVNSNMNMTLNMNSYRTMQPRMANSYNSNPAYVMTQAYPM 1951  
 QY 1352 DINGMAQNGNSMFSQ--QSPPHQEQQANTSYNSNMN 1389  
 DB 1952 QMO---MGMMGSOAYTQOQPMQPNBGMMYTGCPSHHSYMN 1988  
 RESULT 29  
 CBP\_MOUSE STANDARD; PRT; 2441 AA.  
 ID CBP\_MOUSE  
 AC P45481;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE CREB-binding protein.  
 GN CREBBP OR CBP.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=94019866; PubMed=8413673;  
 RA Christia J.C., Kwok R.P.S., Lamb N., Hagiwara M., Montminy M.R.,  
 RT "Phosphorylated CREB binds specifically to the nuclear protein CBP.",  
 RL Nature 365:855-859(1993).  
 CC -FUNCTION: MEDIATES CAMP-GENE REGULATION BY BINDING SPECIFICALLY TO  
 CC PHOSPHORYLATED CREB PROTEIN. ACTING AS A COACTIVATOR, CBP AUGMENTS  
 CC THE ACTIVITY OF PHOSPHORYLATED CREB TO ACTIVATE TRANSCRIPTION OF  
 CC CAMP-RESPONSIVE GENES.

```

CC -1 SUBCELLULAR LOCATION: Nuclear.
CC -1 SIMILARITY: CONTAINS 1 BROMODOMAIN.
CC -1 SIMILARITY: CONTAINS 1 ZF-TYPE ZINC FINGER.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; S66385; AAB28651.1; -
DR TRANSFAC; T01318; -
DR MGD; MGI:1098280; Crebp.
DR InterPro: IPR001487; Bromodomain.
DR InterPro: IPR001487; KIX.
DR InterPro: IPR00197; TAZ_finger.
DR InterPro: IPR000433; Znf_Z2.
DR Pfam; PF00439; Bromodomain; 1.
DR Pfam; PF02172; KIX; 1.
DR Pfam; PF02135; ZF-TAZ; 2.
DR Pfam; PF00569; ZF; 1.
DR PRINTS; PR00503; BROMODOMAIN.
DR SMART; SM00297; BROMO; 1.
DR SMART; SM00291; Znf_Z2; 1.
DR PROSITE; PS00633; BROMODOMAIN_1; 1.
DR PROSITE; PS50014; BROMODOMAIN_2; 1.
DR PROSITE; PS01357; ZF-Z2_1; 1.
DR PROSITE; PS50135; ZF-Z2_2; 1.
DR Transcription regulation; Nuclear protein; Activator; Bromodomain;
KW Zinc-finger.
FT DOMAIN 1104 1176 BROMODOMAIN.
FT ZN_FING 1702 1745 ZF-TYPE.
FT DOMAIN 1062 1065 POLY-GLU.
FT DOMAIN 1556 1563 POLY-GLU.
FT DOMAIN 1944 1949 POLY-PRO.
FT DOMAIN 1968 1971 POLY-GLN.
FT DOMAIN 2082 2086 POLY-GLN.
FT DOMAIN 2200 2216 POLY-GLN.
FT DOMAIN 2296 2299 POLY-GLN.
SO SEQUENCE 2441 AA; 265474 MW; 0AB8028C3112F419 CRC64;

Query Match 3.3%; Score 252; DB 1; Length 2441;
Best Local Similarity 20.6%; Pred. No. 0.00053;
Matches 248; Conservative 142; Mismatches 442; Indels 372; Gaps 61;

QY 392 GKPLNP-----ISSNSPAHQALCSGNPGQDMTLSSNINFPINGPKREOMGMGRFGSGGM 447
DB 78 GSSINRPGIGNVASASSPVQGL--GGQAQGGQPNSTN-----MASLGAM-GK 119
QY 448 NHVSOMQATTPOGSNTALKMNSPSSQSPGMNGOPTSM--LSFRHKMSPCVACSPRIAPSQ 506
DB 120 SPLNGDSSSTP-----NLPKQAASSTGPTPPASQALNPQAOKQVGLVTS----- 163
QY 507 FSPAGSLSPVGVCSSTGNSHSY-----TNSSLNLQALSEGHSVLSGLSSLPDKMGN 561
DB 164 -SPATISQTPG- GICMNAFNPQTHPGLINSNSGSHLMOAQOQAOVYNGSL----- 212
QY 562 LQNSFVNNPPPLSKMGLSKDCFGLYGEPSGTTGQAESSCHPGEOKETNDPNLPPAV 621
DB 213 -----GAAGRGRCGAGMPYPA-----PAMQCAT 234
QY 622 SSERAD-----GOSRLHDSK--GQTKLLQTLITKSDMEESPPLASSLSIDNKDST 669
DB 235 SSVLAETLTQVSPQAMGAGLNTAAGGTTKMGMTGTT-----SPFGQPSQTGGQOM 287
QY 670 GSLPSSGTHGTSLEKHKILHRLQDSSSPVDLAKLTAEATGKDLSDQSSSTAPGSEVT 729
DB 288 GA-----TGVNPQLASKOSMNSL---PAFPDIT-KNITSVTTPVPMNSQLQTSVGIVPTQA 338
QY 730 IKQEPVSPKKENALLRYLLDKDTRDIGLEIPTRKLERLDSKTDPSANTKLIAMKTEKE 789

```

```

DB 339 IATGPFADEKRRKLLQOQLV-----LLJHAHKQRR 369
QY 790 EMSFERGDDQGESELDN---LEELIDLONSQLPQLFPDTRPGAPAGSVKQALINDMLQ 846
DB 370 EDA--NGEVRACSLPRCPTKKNVLYNMHTQCAP-----KACQVAHCASSRQIISHKNC 421
QY 847 TSENSVPRVGAQKALRISQSTFNNRPGQLGRLLPNQNLPLDITLQSPGACPPPIR 906
DB 422 TTHDCV-----CLPKNADSKRNRQITLG----- 446
QY 907 NSSPYSVIPQPMGMNGQIGNOGLGNSSTGMIGNSASRPMPSPGEMAPQSSAVRVTCA 966
DB 447 --SPAS-----GIQNTIGSVGAGQGNATSL--SNBNPTDP-----SSMRAVYA 486
QY 967 ATTSAMNRPVQGMIRNPAASIIMRPSQPGQRQT--LQSQVYN--TGPSELEMMGPGQYS 1024
DB 487 LCLPYKNOP-----OTQLOPYVPGQOPAPQAPRAHQKRTLALGNNMVASVPAGITTD 538
QY 1025 QOQAPNPQTPAPWPESTILPIDQSFASQNRQPF---GSSPDDL--LCPRPAESPDEGA- 1078
DB 539 QQ--PPNLLS---ESALP---TSLGATN--PLMNGSNGNIGSLSTIPTAAPSSGTGR 588
QY 1079 -----LLDQLYLAL-RNFDGLEETIDRALGIPELVQSQAADPQFSSQDSN 1123
DB 589 KGMHEHVTDLRSHLVKHLVQAIFPTPDPAALKDRM--ENLVAAVKVEGDMYESANSR 646
QY 1124 ----IMLEQKAVFPOQVYSAQMAQGSYPMQDPNFHTMGORPSATYLRMPRRLART 1179
DB 647 DEYVHLLAKET-----YKIOKELEEKRRTRLRKQGI--LGNOPALPASGAOP-PVITPPA 697
QY 1180 GLVONOPNQLRLQLOHRLQAOONROPLMNQISNVSNVNTLNR----- 1221
DB 698 QSVRRPNGLPLRPV--NRQVSOG-----MNSFPMISGIVQLQAPRMGPAASPMHNSVM 752
QY 1222 -----PGV---PTQAPINAOMLAORREILNOLRORMOQOQVQRTLMNR-----GQ 1268
DB 753 NSMASVPMGAISSRMPQPPNNMGTHANNIMAAQPLQNFPLQNPQSSSGAMSVSVGM 812
QY 1269 GLMTPRSMVAPSGMP-ATMSNP---RIPOANQOQPFPPNYTGISOOP----- 1311
DB 813 GQPAAGVSOQOEPAALPNPLMLAPQAS--QLPCRP---VTQSLPLRTPPPASTAG 867
QY 1312 -----DPGFT-----GATTPQSPILMS-----PRMAYHQS--PMQO---QSOA 1343
DB 868 MSLDHPTRAPGMPTRPQAPARTPSTPVSQGTPTRPGSVPSAAQTQSTPTVQAAQAQV 927
QY 1344 NPAYQAPSDINDGMAQNMGNMFSQSP--PARGQOANTSMYNNNNIN---VSNATNTG 1399
DB 928 TPQPTP-----VQPPSVATPQSSQOQPTPVHTQPGTRPLSQAAASIDNRVPTSTVRS 981
QY 1400 GMSNMQMTGQISMTSV--TSVSTSGILSMGRQVANDPALRGNLPLPNQLRGMDMKQED 1458
DB 982 APTSSQPPPDVPMLEMKTEVQTDAPRPTESKGPR--SEMMEEDLOGSSQVKEED 1038
QY 1459 TTRK 1462
DB 1039 TTEQ 1042

RESULT 30
YAVL_SCHPO STANDARD: PRT: 1794 AA.
AC Q10172:
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Hypothetical 193.3 kDa protein C27F1.01C in chromosome I.
GN SPAC27F1.01C OR SPAC25G10.09C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.

```











```

CC LIGHT RESPONSE.
CC - POLYMORPHISM: THE VARIABILITY IN LENGTH OF THE POLYGLUTAMINE
CC STRETCH IS DUE TO POLYMORPHISM OF THIS REGION.
CC - SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
CC TRANSCRIPTION FACTORS.
CC - SIMILARITY: CONTAINS 2 PAS (PER-ARNT-SIM) DIMERIZATION DOMAINS.
CC - SIMILARITY: CONTAINS 1 PAS-ASSOCIATED C-TERMINAL (PAC) DOMAIN.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: AF067207; AAD10630.1; -.
DR EMBL: AF065133; AAC39101.1; -.
DR EMBL: AF069997; AAC62234.1; -.
DR FLYBASE: FBgn0023076; CLK.
DR InterPro: IPR003015; HLH_Myc.
DR InterPro: IPR001092; HLH_dim.
DR InterPro: IPR001067; Nucleosylator.
DR InterPro: IPR001610; PAC.
DR InterPro: IPR000014; PAS.
DR Pfam: PF00010; HLH; 1.
DR Pfam: PF00785; PAC; 1.
DR Pfam: PF00989; PAS; 2.
DR PRINTS: PR00785; NCTRNLOCATR.
DR SMART: SM00353; HLH; 1.
DR SMART: SM00086; PAC; 1.
DR SMART: SM00091; PAS; 2.
DR PROSITE: PS00038; HELIX_LOOP_HELIX; 1.
DR PROSITE: PS50112; PAS; 1.
KW Transcription regulation; Nuclear protein; Repeat; Biological rhythms;
KW DNA-binding; Alternative splicing.
FT DNA_BIND 12 24 BASIC DOMAIN.
FT DOMAIN 25 62 HELIX-LOOP-HELIX MOTIF.
FT DOMAIN 84 154 PAS 1.
FT DOMAIN 251 317 PAS 2.
FT DOMAIN 548 559 POLY-GLN.
FT DOMAIN 766 769 POLY-GLN.
FT DOMAIN 794 836 POLY-GLN.
FT DOMAIN 874 877 POLY-ASN.
FT DOMAIN 887 895 POLY-ASN.
FT DOMAIN 953 963 POLY-GLN.
FT DOMAIN 1023 1023 IMPLICATED IN THE CIRCADIAN RHYTHMICITY.
FT VARIANT 816 823 MISSING.
FT CONFLICT 12 12 K->KSFIC (IN REF. 3).
FT CONFLICT 32 32 N->D (IN REF. 3).
FT CONFLICT 128 128 N->K (IN REF. 2).
FT CONFLICT 555 555 N->S (IN REF. 1).
FT CONFLICT 605 605 I->L (IN REF. 3).
FT CONFLICT 912 912 Y->C (IN REF. 3).
SO SQUENCE 1023 AA; 115751 MW; 51437ACBC050DAFB CRC64;

```

Query Match 3.3%; Score 249; DB 1; Length 1023;  
 Best Local Similarity 17.7%; Pred. No. 0.00024;  
 Matches 241; Conservative 176; Mismatches 484; Indels 458; Gaps 49;

```

QY 4 MGEHTSPSRAETRRKKECDQQLSPSPKRN-TEKRNREQENKTYEELAELEIFANFNIDN 62
   | : : : | : : | : : | : : | : : | : : | : : | : : | : : | : : |
Db 1 MDESDD----KDDTKRK-----SRNISEKKRRDQFNLSLVNDLSALISTLS----- 41
QY 63 FNFPRDCAITIKETVKQIDROIKRQEKAAANIDEVQSDVSSTQGVITDADALGPMLEA 122
   | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : |
Db 42 -SRKMDSTVLKSTIALK--NNNEATDRSKVFETIQ---DMKPAFLSNDETHYTHLES 94
QY 123 LDGFEFVNLEGANVFSENVQYLRVNOEELMKSVYSILHVGDTHEFVKNL--PKSI 180
   |||| | : : | : : | : : | : : | : : | : : | : : | : : | : : |
Db 95 LDGEMVVFSSMGSTFYASESITSQLGTLPLDLYMTIYDLAYEMDH--EALLINIMNPTPV 153

```

```

QY 181 VNGGWSGDEPRR-----NSHFF-----NCRMLVK-----PLPD 209
   | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : |
Db 154 IE-----PRODSSNQDITFEYTHLRGGMKGVADANAYELKPFQGYFRNDTNTSGS 205
QY 210 SEE-----ESHQNGEAKHKEYETQNCQSAVGPQSKIKREGEDLQSLCVA 253
   | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : |
Db 206 SSEVSNSSNGCPAVALPRIFQONPAEYDKKLVFGTGRVQDPOLIRMS----- 254
QY 254 RRVKMRKRPVLPSSSEFTRODLQKITSIDTSTPMRAAMPGWEDLVRCLOKFAHQEG 313
   | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : |
Db 255 -----IIDPTNETSKHSMEMKFLFDHRAAPRIIGYMFVVLQTSQDYDHHFDLD 306
QY 314 ESYVAKRHHEVLRQ--GLAFSQYLRFSLSLDTLVAQTKSLRISQTNBPOLVLSHM 372
   | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : |
Db 307 SIYAC-----HEELRQTEGEGSCYRFLTGQOQIMWLQTYVYSYHOFNSKPRDYVCTHK 361
QY 373 LHRQNVYVNNPDLTGQTMCKPLNPISNSPAHALCSGPNQDMTSSNI--NPPINGPK 431
   | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : |
Db 362 VVSYAEV-LKDSRKREGKSGNS-NSITNNGSSKYIASTGTSKASATVTLRDELS-- 417
QY 432 EOMGMPGRFGSGGGMNHVSGMQATTPOGSNYALKMSPQSSPGMNPQPTSMISPRHR 491
   | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : |
Db 418 -----QNDSTLIGNSLA-----SLGTETA 438
QY 492 MSPGVAGSPRIPSQSPAGSLSPVGVCSSTGNSHYTSSNLALQALSEGHSVLSGSS 551
   | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : |
Db 439 TSPAVDSSPMMSASAVPQSGS----- 459
QY 552 LASPDLKMGNIQNSPVMMNPPPLSKMGSIDSKOCFGLYGE--PSRGTGQAESSHCPQEK 610
   | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : |
Db 460 -----QINPLKTSRASS-----YGNISSTGISPKARKKY----- 491
QY 611 ETNDPNLPAVSSERRADQSGSLHDSKQTKLIDLITTKSQDMERPSSPLASSIDPNKDSG 670
   | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : |
Db 492 -----FYNNRQND-----SDSTMSSTD 508
QY 671 SLPGSGS--THGTSLKAKHKILRLDODSSPVLDLAKTLAEATGKDLQSSSTAPGSEV 728
   | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : |
Db 509 SVTSRQSMTHVSSOSORSHNR-----EHNRENHNHNSH 545
QY 729 TIKREPSPKKENALLRYLLDKDQTDIGLRETPPLERLDSKTPDASPMTKLAMTEK 788
   | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : |
Db 546 HMQDQOQHONQOQHONQOQHONQOQHONQOQHONQOQHONQOQHONQOQHONQOQHON 595
QY 789 EEMSFERDQPGSELMLIEILLDLQNSQLPQ--LEPDRTRG--APAGSVKQALINDMQ 845
   | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : |
Db 596 -----CQFPQAVPLASQVLAAPVLEPPQYLTALPMQ 628
QY 846 LTAENSVYTV---GAQKTALRISQSTFNNPRGQLG-----R 880
   | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : |
Db 629 PVIAFPFVAVLSPRLPVQSQTDMLRDTVVMTPQSQDQDQDQDQDQDQDQDQDQDQD 688
QY 881 LLRNQNLRLDITLQSPGAGFPFRIRNSSPYVLRQGMGNQ--GMIGNQNLGNSSTG 938
   | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : |
Db 689 IVSEQLLSRTYTLQPMMSGFAP-----GNMTAAVGNIGASGQGLN 732
QY 939 MIGNSASRPTMPSEGMAPOSSAVRVTCATTSAMNRPVQGMIRPAASJPMRSPSQGQ 998
   | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : |
Db 733 FTGGSNAVQPOFNGQGFALNSEQM-----LNQDQDQMMQ-----Q 767
QY 999 KQTLQSOVMNIGSELEMMNGGPOYSQOQAPRPNQTAWPSPSILPDIQASRASQNRQFQS 1058
   | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : |
Db 768 QONLHTQHNHLQOQHOSHQDQHTQOQHONQOQOQOQOQOQOQOQOQOQOQOQOQOQO 823
QY 1059 SPDDLCPHRAESPDEGALLDQYLALNPFQSELEIDALGIREL--VSQSAVAPDEQ 1116
   | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : |
Db 824 -----QOQOQOQDQDQOQNDITLLR--EDIDIDAFNLSPHLSIGOSTTIPFN 870
QY 1117 FSSQSNIMELOKAPVPPQOYASQQAQAG-----SYSPMDPPNFTNGQRPYATLR 1169
   | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : |
Db 871 SSSNNNN-----QSYNGSSNLNNGNQNNNNNSSNPQNNNDSL--LSYQMA 916
QY 1170 MQPRP-----GLRPTGL--VQNOPNQLR---LQLQHLQAOQNRQPLMQNISVSNVL 1218

```



QY 821 L--FP 823  
DB 782 LTMFP 786

RESULT 35  
HMCU\_DROME  
ID HMCU\_DROME STANDARD; PRT: 2175 AA.  
AC P10180: 09W306:  
DT 01-MAR-1989 (Rel. 10, Created)  
DT 01-MAR-1989 (Rel. 10, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Homeobox protein cut.  
GN CT OR CG11387.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=88232956; PubMed=2897632;  
RA Blochlinger K., Bodmer R., Jack J., Jan L.Y., Jan Y.N.;  
RT Primary structure and expression of a product from cut, a locus  
RL involved in specifying sensory organ identity in Drosophila.";  
RL Nature 333:629-635(1988).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BERKELEY;  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
RA Baller R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,  
de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Foslcr C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostlin D., Houston K.A., Howland T.J., Wei M.-H., Idegawa C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., Moberg D.,  
RA Merkulov G., Malsbina N.V., Moberg C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclab J.R.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E.C., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of Drosophila melanogaster.";  
RL Science 287:2185-2195(2000).  
CC -1- FUNCTION: REGULATOR OF CELL FATE DECISIONS IN MULTIPLE LINEAGES.  
CC SPECIFICALLY, FUNCTIONS AS A DETERMINATION FACTOR THAT SPECIFIES  
CC SENSORY ORGAN IDENTITY IN PRECURSOR CELLS. PROBABLY ALSO INVOLVED  
CC IN CELL TYPE SPECIFICATION OF MALPIGHIAN TUBULES. IN ABSENCE OF

CC CUT GENE EXTERNAL SENSORY ORGANS ARE TRANSFORMED INTO CHORDOTONAL  
CC ORGANS.  
CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).  
CC -1- TISSUE SPECIFICITY: DETECTED IN MANY CELLS IN THE CENTRAL NERVOUS  
CC SYSTEM, ALL EXTERNAL SENSORY ORGANS, SOME PERIPHERAL NEURONS, AND  
CC IN THE NON-NEURAL CELLS OF THE SPIRACLES AND THE MALPIGHIAN  
CC TUBULES.  
CC -1- DEVELOPMENTAL STAGE: CELL-SPECIFIC PATTERN OF EXPRESSION. BROADLY  
CC EXPRESSED DURING EMBRYONIC DEVELOPMENT.  
CC -1- DOMAIN: ASN AT POSITION 47 OF THE HOMEBOX MAY PARTICIPATE IN  
CC REGULATING DNA-BINDING ACTIVITY BY PROMOTING HOMO- AND  
CC HETERO-DIMERIZATION.  
CC -1- SIMILARITY: CONTAINS 3 CUT DOMAINS.  
CC -1- SIMILARITY: BELONGS TO THE CUT FAMILY OF HOMEBOX PROTEINS.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: X07985; CA30794.1; -.  
CC EMBL: AE003441; AAF46264.2; -.  
CC PIR: S03170; S03170.  
CC HSSP: P04002; IWFA.  
CC TRANSFAC: T02004; -.  
CC FlyBase: FBgn0004198; ct.  
CC InterPro: IPR003350; CUT.  
CC InterPro: IPR000047; HTH\_repressor.  
CC InterPro: IPR001356; Homeobox.  
CC Pfam: PF002376; CUT; 3.  
CC Pfam: PF00046; homeobox; 1.  
CC PRINTS: PR00031; HTHREPRESSR.  
CC SMART: SM00389; HOX; 1.  
CC PROSITE: PS00027; HOMEBOX\_1; 1.  
CC PROSITE: PS50071; HOMEBOX\_2; 1.  
CC Transcription regulation: Homeobox; DNA-binding;  
CC Developmental protein; Nuclear protein; Repeat; Coiled coil.  
FT DOMAIN 265 343  
FT COILED COIL (POTENTIAL).  
FT DOMAIN 433 499  
FT COILED COIL (POTENTIAL).  
FT DNA\_BIND 798 964  
FT COILED COIL (POTENTIAL).  
FT DOMAIN 1056 1161  
FT COILED COIL (POTENTIAL).  
FT DNA\_BIND 1329 1417  
FT COILED COIL (POTENTIAL).  
FT DOMAIN 1463 1522  
FT COILED COIL (POTENTIAL).  
FT DNA\_BIND 1608 1695  
FT COILED COIL (POTENTIAL).  
FT DOMAIN 1745 1804  
FT COILED COIL (POTENTIAL).  
FT DOMAIN 194 210  
FT COILED COIL (POTENTIAL).  
FT DOMAIN 235 243  
FT COILED COIL (POTENTIAL).  
FT DOMAIN 271 293  
FT COILED COIL (POTENTIAL).  
FT DOMAIN 384 428  
FT COILED COIL (POTENTIAL).  
FT DOMAIN 547 554  
FT COILED COIL (POTENTIAL).  
FT DOMAIN 574 584  
FT COILED COIL (POTENTIAL).  
FT DOMAIN 616 630  
FT COILED COIL (POTENTIAL).  
FT DOMAIN 665 699  
FT COILED COIL (POTENTIAL).  
FT DOMAIN 2004 2014  
FT COILED COIL (POTENTIAL).  
FT DOMAIN 2071 2077  
FT COILED COIL (POTENTIAL).  
FT DOMAIN 2124 2136  
FT COILED COIL (POTENTIAL).  
SQ SEQUENCE 2175 AA; 233628 MW; 08BF00C4861BD0AB CRC64;

Query Match 3.2%; Score 248; DB 1; Length 2175;  
Best Local Similarity 19.2%; Pred. No. 0.00069;  
Matches 305; Conservative 202; Mismatches 669; Indels 416; Gaps 68;

QY 7 NTSDPSRAETFRKKRCEPDQIGPSFKRNTKRNDEQNKYIEELAEILFANFNIDNENK 66  
DB 394 NTSNTATSNNTNNNN-----NSSSGNSEKRRKNNNNNNQPAVLLAARDKEI----- 442  
DB 67 PDKCALKEIKVQIROIKEQKAAANID-----EVKSDVSTGGVYIDKDALGPMML 120  
DB 443 -----KALLDELQRLRAOEGTHLVQIQLRLEHLEVKRHHIRL-EARLDKQOINEALA 494

```

QY 121 EALDGEFFVNLGNVNFVSENVTOYLRY NOEELMKNKSVSYLHVGDTHEFYKNLLPKSI 180
D 495 EA-----TALSAASATNNNNNSOSSDNKK--LNTFAERMOASSNAD-----LPSS- 539
QY 181 VNGSGSEPRRRNSHTNCRLVPLP---DSEEGH-----DNQANOKYETMOCFAVS 233
D 540 -----TKAPVAEDEDEEDQAMLVDESEADKPED-----S 571
QY 234 OPKSKEGEDLOS-----CLICVARHVPKERVLPSSSEFTTRQDLOGKITSL 283
D 572 HHDDDEDEDEBEAVNATTTTSNELKTKEDHSPDLVLNLSNSAIAAAAAAAAAAACAN 631
QY 284 DSTMRA-----AMKPG-----WEDLYRRCIOKFFHAGESEVSYAKRH----- 322
D 632 DPNKFQALLIBRTKALAEALKNASDALSEDANHHQOQHQQOQHQQOQHQQOQHQQO 691
QY 323 HHEVLROGLASFQYRSLSGTLVAOQTKSLIRSOITNE-----POLVI 368
D 692 HHHHLOQ-----OPNSGNSNPASNNHHGHNLHGHLHPSSAH 731
QY 369 SLHMLHRONYCVNMPDLTGOTMGRPLNPISNSPAHALCSGNGODMTLSSNINPFI 428
D 732 HLHHQTTESNSNSTPTAGANNNGS--NNSSNT-----NANSTAOALAASLASTLN 780
QY 429 GPKEDMGMPGRFGSGGMNHVSGMOATTPQGSNYALKNPSPOSRCMPNFGQPTSMLEP 488
D 781 GTSKLM-----QEDSNGLAAYA-----MAHAHQAAALGPFGLGRFQFRA 823
QY 489 RHRMSPGVA-GSPRIIPS--QFSPAGSLHSPVVC-----SSTGNSHYTN 531
D 824 AQAAGGGRGHYRFADSELDLPFGASMAGRLESLPKDPMKAEKLEMLRYNNMDKYAN 883
QY 532 SSLNAL-----QALSEBGHVSIGSSLASP---DLKNGNLONSPYNNMPPRLSKGSDS 582
D 884 QALDTLHSRVRELVSYH--NIGQRFKAKYILGISOQTV--SELLSKPKWMDKTEKGR 939
QY 583 KDCFGLYGEPEEGTGAQESSCHPEQKETNDPLRPVASSERADGOSRLDOKGOTKL 642
D 940 DSYKMHMAMACDDNAVMMLKSLIP-----KKGSGLPYAG--RAGGAGGSDSMEDEITA 992
QY 643 QLLTTKSDQMEPSPLASL-----SDTKNDSTGSLPGSGSTHGTSLKEKHKILH 691
D 993 HILSEASSILMKOSSVACQHREOERSHGEGDSHNSDEKS--PQSGTSPPFEVENOLKH 1050
QY 692 RLDDSSSPVOLAKTLATGATKDLDSQESSSTAPSEVITIKOEPVPRKKENALLRYLLDK 751
D 1051 QHLNPEQAAAOQEREREQEREQOQRLRHDDQDMARLYOELLARTPRETAFPSFL-- 1107
QY 752 DDTKDIGLPEITPKLERLDSKTDPAASNKLIAKTEKEEMSFEPDQDQSGEIDNLEETLD 811
D 1108 -----FSPSLFGGAACMPGASNAFPAMADENNRHFE-----RELAKIOQOHO 1151
QY 812 DLONSQPLPFPTRPGAPAGSVDKOAIINDL--MQLTAENSPVTYVGAOKTALRISOST 869
D 1152 QOQAQAQAOQRPN-----FSSLALAOQOVLNGAODLSIAAAAKDKIKLQSSSLEHSAGS 1207
QY 870 FNNRPGOLGRLPLNONPLDITL---OSPAGGPEPPIRMS-----SPYSVIRPOGM 919
D 1208 SSCCKOBERDAYPS-----SLHGKKSBEGGGTRPAPRPPSGRETGAGAPPTAAPTG- 1259
QY 920 MGNOMGIMNOGLNLSGTMGISASRPTMPS--GEMAPQSSAVRYTCAATTSAMNRRVQ 977
D 1260 -----GASNN--SAAPSPLSNSTILPPALSSQGEFFAATASPIQRMASINSLSLITQRPV 1310
QY 978 GGMTRNPAASLTPMRPSSO--PGQROQLOSVMNIGPSEL--EMNNGGQOYSGOQAPPNQT 1033
D 1311 -----TPHNSTPORPRTKAVLPPITQOQFDMFNNTATEDIVRRVKEALSQOYISIQ----- 1359
QY 1034 APWPESTLPIDQASFASQNRQPFSSPPDILCPRHAAESPDEGALLDQL-----YLALRN 1089
D 1360 RLFBESVYIGLSQ-----GSVSDLLARPKP-----WHMLTOKGREPIIRKMK 1400

```

```

QY 1090 FDGLEIDRALGIPELVVSQAVDEDEQFSODSNITMEOKARVPFPOYASQAOMAGSYS 1149
D 1401 F--LEDEN--AVHKLVASQYKIAPEKLMRTGS-----YSGSPQMRQGLASKMQAAS--L 1448
QY 1150 PMQDPNFTMGORSEYATLRLMQPRGLRPTGIVQNPOLRLQDOLHRLQAOQNRPLMNO 1209
D 1449 PMQKMSSELKLOEPRAOAHLMQOAMASAMQOO--QVAAQOQAOQAOQAOQAOHLOOQ 1506
QY 1210 ISNVSNVULLTRGVPTQAPINAOMLAORERILMQLRORMHQOQOQVQORTLLMRGQ 1269
D 1507 -----AQOHLQOQ-----QHLAQOQ--HPHQOHHQ----- 1529
QY 1270 LNMTPSNVAPSGPMATMSPNPRIPQANAOQFPFPYNGISQOQPDGFT--GATTP--OSP 1324
D 1530 --AAAAAALHHQSMLLNSPGLRPQHA--ISLPPGAGGAQPGCGPGNGSSNPSNSEKPP 1585
QY 1325 LMSP-----RMAHT--OSPMMQOSQANPAYOAPSDIN----- 1354
D 1586 MLMPHGTNAMSRLQOHSMPVYEEMALTOPDLDTITTKIKEALLANNIGOKIRGEAVL 1645
QY 1355 GMAOENMGNSMFSQOSPFPF-----GOQANTSMY-----SNMNMNINVMATITGMSMN 1405
D 1646 GLSOGSV--SELLSKPKFWMHLSIKGREPFIROMQLMSDANNVERLQLLKNEREASRR 1703
QY 1406 QMTG-----OISMITSVTSTSGLSMGPEDQY 1432
D 1704 RSTGPNQODNSDTSNSDTNDEFTYSRPGGSV 1735

```

```

RESULT 36
TRX_DROVI STANDARD; PRT: 3828 AA.
AC 024742:
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Trithorax protein.
GN
OS Drosophila virilis (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7244;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96100387; PubMed=8555104;
RA Tiliid S., Sedkov Y., Mizrokh L., Mazo A.;
RT "Conservation of structure and expression of the trithorax gene
between Drosophila virilis and Drosophila melanogaster.";
RL Mech. Dev. 53:113-122(1995).
CC -!- FUNCTION: FUNCTIONS IN SEGMENT DETERMINATION THROUGH INTERACTION
WITH GENES OF TRITHORAX (BX-C) AND ANTENNAPEDIA (ANT-X) COMPLEXES.
CC IF CAN BEHAVE AS AN ACTIVATOR OF BX-C.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: BELONGS TO THE TRANSCRIPTION FACTOR TRITHORAX FAMILY.
CC -!- SIMILARITY: CONTAINS 1 'SET' DOMAIN.
CC -!- SIMILARITY: CONTAINS 5 PHD-TYPE ZINC FINGERS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
DR EMBL; Z50038; CA90349.1; -.
DR HSSP; P19793; 2NIL.
DR FlyBase; FBgn0014844; Dvir\trx.
DR InterPro; IPR003889; Flyrich_C.
DR InterPro; IPR003888; Flyrich_N.
DR InterPro; IPR001965; PHD.
DR InterPro; IPR003616; PostSET.

```



OS Drosophila melanogaster (fruit fly).  
 CC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 CC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 CC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RP [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Embryo;  
 RX MEDLINE=9613612; PubMed=8557198;  
 RT Wilk R., Weizman I., Shilo B.-Z.;  
 RT "Trachealless encodes a bHLH-PAS protein that is an inducer of  
 RT tracheal cell fates in Drosophila.";  
 RL Genes Dev. 10:93-102(1996).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Embryo;  
 RX MEDLINE=9613613; PubMed=8557189;  
 RA Isaac D.D., Andrew D.J.;  
 RT "Tubulogenesis in Drosophila: a requirement for the trachealless gene  
 RT product.";  
 RL Genes Dev. 10:103-117(1996).  
 RN [3]  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RC STRAIN=BERKELEY;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle J.D.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blaise R.G., Champs E., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Barker E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abell J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Bernier B.P., Bhandari D., Bolintiner P.,  
 RA Botkova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,  
 RA Butlis K.C., Busam D.A., Butler H., Cadieu L.B., Cantler A., Chandra I.,  
 RA Chertys J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrier S., Fleischmann W.,  
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodet A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Maltel B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclab J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J.D., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Klimos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zhang X.H., Zhong F.N., Zhou W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster.";  
 RL Science 287:2185-2195(2000).  
 CC -1- FUNCTION: TRANSCRIPTION FACTOR, MASTER REGULATOR OF TRACHEAL CELL  
 CC FATES IN THE EMBRYO. NECESSARY FOR THE DEVELOPMENT OF THE SALIVARY  
 CC GLAND DUCT AND THE POSTERIOR SPIRACLES. IT MAY INDUCE A GENERAL  
 CC FATE OF BRANCHED TUBULAR STRUCTURES OF EPITHELIAL ORIGIN. TGO/TRH  
 CC HETERODIMERS ARE INVOLVED IN THE CONTROL OF BRACHYLESS EXPRESSION.  
 CC -1- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER  
 CC BHLH PROTEIN. HETERODIMER WITH TGO.  
 CC -1- SUBCELLULAR LOCATION: Nuclear.  
 CC -1- ALTERNATIVE PRODUCTS: AT LEAST 3 ISOFORMS; 1 (SHOWN HERE), 2 AND  
 CC 3; ARE PRODUCED BY ALTERNATIVE SPLICING.

CC -1- TISSUE SPECIFICITY: TRACHEA, SALIVARY GLAND DUCTS, POSTERIOR  
 CC SPIRACLES (FLITZKOEPER PRIMORDIA) AND A SUBSET OF CELLS IN THE CNS.  
 CC -1- DEVELOPMENTAL STAGE: DURING EMBRYOGENESIS, FIRST DETECTED IN THE  
 CC TRACHEAL PLACODES AT STAGE 8, AND EXPRESSION CONTINUES THROUGHOUT  
 CC EMBRYONIC AND LARVAL DEVELOPMENT. IN THE DEVELOPING SALIVARY  
 CC GLAND, EXPRESSION IS OBSERVED IN THE ENTIRE GLAND AT STAGE 9 AND  
 CC BY STAGE 12, EXPRESSION IS CONFINED TO THE SALIVARY DUCTS.  
 CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF  
 CC TRANSCRIPTION FACTORS.  
 CC -1- SIMILARITY: CONSTAINS 2 PAS (PER-ARNT-SIM) DIMERIZATION DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 1 PAS-ASSOCIATED C-TERMINAL (PAC) DOMAIN.  
 CC -1- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO FRAMESHIFTS  
 CC IN POSITIONS 22 TO 34.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib.ch](mailto:license@sib.ch)).  
 CC -----  
 CC DR EMBL: U33427; AAA96257.1; ALT\_INIR.  
 CC DR EMBL: U42699; AAA96754.1; ALT\_FRAME.  
 CC DR EMBL: AE003468; AAF47386.1; -.  
 CC DR Flybase: FBgn003749; trh.  
 CC DR InterPro: IPR003015; HLH\_Myc.  
 CC DR InterPro: IPR001092; HLH\_dlm.  
 CC DR InterPro: IPR001610; PAC.  
 CC DR InterPro: IPR000014; PAS.  
 CC DR Pfam: PF00785; PAS; 1.  
 CC DR Pfam: PF00989; PAS; 2.  
 CC DR SMART: SM00353; HLH; 1.  
 CC DR SMART: SM00086; PAC; 1.  
 CC DR SMART: SM00091; PAS; 2.  
 CC DR PROSITE: PS00038; HELIX\_LOOP\_HELIX; 1.  
 CC DR PROSITE: PS0112; PAS; 2.  
 CC DR Developmental protein; Nuclear protein; Transcription regulation;  
 CC Repeat; DNA-binding; Alternative splicing.  
 CC KW DNA\_BIND; 86  
 CC FT DOMAIN 100 140  
 CC FT DOMAIN 174 244  
 CC FT DOMAIN 391 461  
 CC FT DOMAIN 465 508  
 CC FT DOMAIN 629 636  
 CC FT DOMAIN 154 157  
 CC FT DOMAIN 244 250  
 CC FT DOMAIN 251 254  
 CC FT DOMAIN 255 264  
 CC FT DOMAIN 722 728  
 CC FT VARSPLIC 281 286  
 CC FT VARSPLIC 328 356  
 CC FT VARSPLIC 78 78  
 CC FT CONFLICT 250 250  
 CC FT CONFLICT 703 703  
 CC FT CONFLICT 708 708  
 CC FT CONFLICT 829 829  
 CC FT SEQUENCE 958 AA; 102241 MW; 8F9CF758F1370541 CRC64;  
 CC  
 CC Query Match 3.2%; Score 243.5; DB 1; Length 958;  
 CC Best Local Similarity 18.0%; Pred. No. 0.00039;  
 CC Matches 207; Conservative 152; Mismatches 361; Indels 433; Gaps 50.  
 CC  
 CC OY 10 DSRATRRKKKCPDGLGSPKRNTEKRNRENKYLEELAFANFNIDNFKP-- 67  
 CC : | | | | : | : | | | | : |  
 CC DB 80 EGIILELRKK-----SRDARSRRGKENEFYELAKML-----PLP 116  
 CC : | | | | : | : | | | | : |  
 CC OY 68 -----DKCALIKETVKOI-----KQIKQEKAAANIDVOKSDVSSTG 106  
 CC : | | | | : | : | | | | : |  
 CC DB 117 AATISQDASIIIRLTISYLKLRDPSGHDPWTRASSSSKIKSAIRSPAYDLFEQH 176  
 CC : | | | | : | : | | | | : |  
 CC OY 107 GCVTDKDALGPMALDGLFFVNVLEGNVVESENVTYLRYNGDELMANKSVYSILHWG 166  
 CC : | | | | : | : | | | | : |

```

Db 177 QGT-----HIIQSLDGFALAAADGFLYISVYITLGLSQVEMTSSITFDYTHQA 228
QY 167 DHTFVKNLLPKSIVNGSGSNG-----EPPRRNSHTFNCRL 203
Db 229 DHSF-INDQJGLSTSGGGGGSSSSGGGGGAGGAGMASPTSGASDGGSTHCTNNPDV 287
QY 204 VKPLPDSSEEGHNOF-----AHQYE-----TWOC-----FAVSQPKS 237
Db 288 AASTQTASTGSKGYDRSFCVRMKSTLTJKRCHEKSSGYRASDATSCNNGNNAASNNAKN 347
QY 238 IKKEGELQGLICVARRV-----PKMERPVL-----PS-----SESFTT 272
Db 348 VKNPGSNVSVLLCKLRPQYTTSHSRKSOPLGLGVALAIALPPPSVHEIRECDMFTV 407
QY 273 RODLOGKITS-----LDTSTWRAAMKPGMEDLVRCIOKF-HAQHEGESVYAKRH 323
Db 408 RVNFDLRAVANCEPRVSDLLDYS-----EDLVNKSILSLCHA-----EDANRLAKSH 454
QY 324 HEVLRQGLAFSQYRFSLSQGLVAAQTCKSLIRSOTNEPQVLSLHML-HREQ----- 377
Db 455 SDLEKGOVLTYGYRLNKKSGGYTWLQTCATVVCSTKNADEQNIICVNYVISNRENM 514
QY 378 -NVCVMNPDLTGQTMKRLPISNSPAHQ-----LCSGNPGQDMTLSSINPPIGPK 431
Db 515 LDCCQLRP-----SPDSIKHEEGLGNDKSSSGSGD----- 545
QY 432 EOMGMPMGREFGSG-GMNHVSGMOATTPQGSNVAALKMNSQSSSPGNNPQPTSMLSPRH 490
Db 546 -----ASGEKSHLSAGD-----MKLNSPTDEG----- 570
QY 491 RMBPGVAGSPRIPPSGFSPAGSLHSPVGCSSGTGNSHSYTNSSALNALQALSEGH---GVS 547
Db 571 -----HSHRGGRSAASH---GSSMNSLTMLKIDSPPTPLGVE 604
QY 548 LGSSASPDLKMGNLQNSPVNMNPPRLSKMGSLSKOCFGLYGPSPGCTGQAASSCH- 605
Db 605 IDSCVLEPTTVA-----TPPAATPVQ-----STKRKRKTASOAE 641
QY 606 -PGOKETNDPNLPPAVSSERADGOSRLHDSKQTKLLOLITTKSDMERSPLASLSDT 664
Db 642 DQGEQVYSQPLPKPLTMQRQOQPR-----SRLPSTV-----DEGESSMADS----- 685
QY 665 NKDSTGSLPGSGSTHGTSLKEKIKLHRLLODSSSPVDLAKLAETATGKDLQSSSSTAP 724
Db 686 -----AVKDLEQAMSKHLPSPAVAVSVAAPNPDFSADSLKQOQ----- 724
QY 725 GSEVTIKOEVSPPKKRENALLRYLDDDKDGLPEITPKLERLDSKTDPA----- 776
Db 725 -----OQOOLDPWEKSTT-----OMIGTPYQQPPAPV-----PATALLRQLY 762
QY 777 SNTKLIAKTEKEEMSFEPDQPSGLNDLEIILDLQNSQLPOLFPDTRGAPAGSVDK 836
Db 763 ANRSVRARARQPT-----GVGPG-----VFYGDQOTGPLP-----TPGGS-----E 801
QY 837 QAITNDLMQJTAENSPLYGAQKTAALRISOSTFNNRPGOLGLFLPNQNLPLDITLQSP 896
Db 802 SSYENQYLIQLSAASGHP-GGOKT-----SADAFTNLVSTYGGY-----HSSIDYHNATPP 853
QY 897 TGAGPPPIPNSSPSYIPOPGMNGMGNOGMLGNSSTGMIGMSASRPTMPSGMAP 956
Db 854 SSVSP-----RDSN-----OPGRKA-----PVLASNGGYIAPPLGLQYAT-----SSGQVVP 897
QY 957 QSSAVRVTCAATTSAMNRPVQGMIRNPASIPMRPSSOPGROTLQSOVANNIGPSELEM 1016
Db 898 -----ATLPLKPOA-----SYATMHPSSGST 919
QY 1017 NMGSPQYSQOQAP 1029
Db 920 TEGGVTYSNLDOP 932

```

```

CBP_HUMAN
ID CBP_HUMAN STANDARD; PRT; 2442 AA.
AC Q92793; Q16376; 000147;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE 01-MAR-2002 (Rel. 41, Last annotation update)
DE CREB-binding protein.
GN CREBBP OR CBP.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97321049; PubMed=9177780;
RA Giles R.H., Petrij F., Dauwerse H.G., den Hollander A.L.,
RA Lushnikova T., van Ommen G.J.B., Goodman R.H., Deaven L.L.,
RA Doggett N.A., Peters D.J.M., Breuning M.H.;
RT "Construction of a 1.2-kb contig surrounding, and molecular analysis
RT of, the human CREB-binding protein (CBP/CREBBP) gene on chromosome
RT 16p13.3."
RL Genomics 42:96-144(1997).
RN [3]
RP SEQUENCE OF 1-405 FROM N.A.
RX MEDLINE=9637968; PubMed=8782817;
RA Borjow J., Stanton V.P., Andresen J.M., Becher R., Behn F.G.,
RA Chaganti R.S.K., Civan C.I., Distèche C., Dube I., Frischauf A.M.,
RA Horstman D., Mielman F., Volinia S., Watmore A.E., Housman D.E.;
RT "The translocation t(8;16)(p11;p13) of acute myeloid leukemia fuses
RT a putative acetyltransferase to the CREB-binding protein."
RL Nat. Genet. 14:33-41(1996).
CC -!- FUNCTION: MEDIATES CAMP-GENE REGULATION BY BINDING SPECIFICALLY TO
CC PHOSPHORYLATED CREB PROTEIN. ACTING AS A COACTIVATOR, CBP AUGMENTS
CC THE ACTIVITY OF PHOSPHORYLATED CREB TO ACTIVATE TRANSCRIPTION OF
CC CAMP-RESPONSIVE GENES.
CC -!- SUBUNIT: INTERACTS WITH SMAD1, SMAD2 AND SMAD3.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- DISEASE: INVOLVED IN ACUTE LEUKEMIAS BY CHROMOSOMAL TRANSLOCATIONS
CC T(8;16)(P11:P13) INVOLVING CBP AND MOZ, AND T(11;16)(Q23:P13.3)
CC INVOLVING CBP AND MLL.
CC -!- DISEASE: DEFECTS IN CREBBP ARE THE CAUSE OF RUBINSTEIN-TAYBI
CC SYNDROME (RSTS), A DISORDER CHARACTERIZED BY CRANIOFACIAL
CC ABNORMALITIES, BROAD THUMBS, BROAD BIG TOES, MENTAL RETARDATION
CC AND A PROPENSITY FOR DEVELOPMENT OF MALIGNANCIES.
CC -!- SIMILARITY: CONTAINS 1 BROMODOMAIN.
CC -!- SIMILARITY: CONTAINS 1 ZZ-TYPE ZINC FINGER.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U47741; AAC51770.1; -
DR EMBL: U85962; AAC51331.1; -
DR EMBL: U89354; AAC51339.1; -
DR EMBL: U89355; AAC51340.1; -
DR MIM: 600140; -
DR InterPro: IPR001487; Bromodomain.
DR InterPro: IPR003101; KIX.
DR InterPro: IPR000197; TAZ_finger.
DR InterPro: IPR000433; ZnF_ZZ.

```





DR Pfam: PF00856; SET: 1.  
 DR Pfam: PF02008; zf-CXXC: 1.  
 DR SMART: SM00297; BROMO: 1.  
 DR SMART: SM00542; FYRC: 1.  
 DR SMART: SM00541; FYRN: 1.  
 DR SMART: SM00249; PHD: 4.  
 DR SMART: SM00508; PostSET: 1.  
 DR SMART: SM00317; SET: 1.  
 DR PROSITE: PS50014; BROMODOMAIN\_2: 1.  
 DR PROSITE: PS50280; SET: 1.  
 DR DNA-binding: Nuclear protein; Zinc-finger; Metal-binding;  
 KW Transcription regulation; Alternative splicing; Polymorphism.  
 FT NON\_TER 1 1  
 FT DNA\_BIND 67 78 A.T HOOK (BY SIMILARITY).  
 FT DNA\_BIND 115 125 A.T HOOK (BY SIMILARITY).  
 FT DNA\_BIND 199 207 A.T HOOK (BY SIMILARITY).  
 FT ZN\_FING 1044 1091 CXXC-TYPE.  
 FT ZN\_FING 1330 1381 PHD-TYPE 1.  
 FT ZN\_FING 1383 1432 PHD-TYPE 2.  
 FT ZN\_FING 1465 1529 PHD-TYPE 3.  
 FT DOMAIN 1605 1650 BROMODOMAIN (DIVERGENT).  
 FT DOMAIN 3737 3846 SET.  
 FT DOMAIN 35 41 POLY-GLY.  
 FT DOMAIN 459 469 PRO-RICH.  
 FT DOMAIN 1231 1238 POLY-PRO.  
 FT DOMAIN 3533 3536 POLY-GLU.  
 FT DOMAIN 3693 3697 POLY-GLU.  
 FT VARIANT 1503 1505 MISSING (IN ISOFORM 2).  
 FT VARIANT 1497 1497 K -> T.  
 SO SEQUENCE 3866 AA; 420976 MW; ADPC55E14E806F1D CMC64;

Query Match 3.2%; Score 243; DB 1; Length 3866;  
 Best Local Similarity 18.3%; Pred. No. 0.0024;  
 Matches 334; Conservative 231; Mismatches 613; Indels 650; Gaps 81;

QY 80 IROIKEDKAAANIDEVOK-----SDVSTGGGVYDKDALGPMLEA-L 123  
 DB 1596 LTEVSKODEQOPDLEGRKMDGYSVYLESPEDITVKIIQAINDGSGQPEIKKNSM 1655  
 QY 124 DGEFFVNLGNGVVESENVTOYLRYNOELMKSVYSILHVGDFEVRN-LIPKSI-V 181  
 DB 1656 VKSEFFIKOMERVPWVSYSKSRFEWPK-----VSNNSMLNANALPESLDH 1702  
 QY 182 NGGSWSGSEPRRNSHTFNCMLYKPLPDESEGHNOEAHQYETMOCFAVSQPKSIKEE 241  
 DB 1703 NYAQW-QEREESHTQPLMKKIIPAPKPGGEDSPPLHPPTLSTDRSEDS 1760  
 QY 242 GE-----DLOSCLIV-----ARRVPMKERP-----VLPSESEFTTRDLO 277  
 DB 1761 PEINPPGIDNRQCALCLMYGDDSDANDAGRLLYIGONETHVNCALMSAEV---EDDD 1817  
 QY 278 GKITSIDTSTMA-----AMKPGME-DLVRRICQFHAQHE-----GESVSYAK 320  
 DB 1818 GSLKNVMAVIRGKOLRCCECQKFGATVGCCLTSCTSNYHPCSRKANCYFLDDKKYYCQ 1877  
 QY 321 RHH-----HEVLRO-----GLAFSQIYRF----- 339  
 DB 1878 RHRDLKGEVVPENGFEVFRVFDPEGISLRKFLNGLEPENHMMIGSTIDCLILN 1937  
 QY 340 SLSDGTLVAQTK-----SKLINSQTTNEPOLVLSLHMLREONVCYANPDLTGOTM 391  
 DB 1938 DLSD-----CEDKLPFGYOCSSRYVSTTDARKKCYCTCKIM-ECRPPVVEPDI-NSTV 1989  
 QY 392 GKPLNPISNSPAH-QALCSGNGQDMTLLSNINPFINPKBGMGMKRFSGSGGMNH 449  
 DB 1990 EHDNKRITIASPSEFIDASKDSOSTAIIIS-----PSEDPHSGTSGSCYYHV 2039  
 QY 450 VSGM-QATTPGGSNYALKMNSPSSGPMNP-----GQPTSMLSBRHR-----MSGV 496  
 DB 2040 ISKVPRIKRTS-----YSPTRSGRCRPLDSAGSPT---PTTHIIVYVGDPLLSGL 2088  
 QY 497 --AGSPRIIPSPQSPAGSLHSPVGCSTGNSHTSNLALQALSEGHVSLGSSLAS 554

DB 2089 RSIGSRHRHSTSLSP-----LNSKLRIMSPVRTGSAYSKSSVSVSPSL---GTATDEPASA 2141  
 QY 555 PDLKMGNIQNSPVNM--NPPL-----SKMGSLDS-----KDCFCGL----- 588  
 DB 2142 KASDRGGLSSSANLGHASAPPPSSSSQRTVGSKTSHLDGSPSEVKRCSALDLVPKGSIV 2201  
 QY 589 YGE-----PSEGTGQAESSCHPEQKET-----NDPNLPAPVAVSE 624  
 DB 2202 KGEKNTRSSKSTGDSASHAYPIKPLIPQVHNAPGELINIKISGFAPRTVP--FSSK 2260  
 QY 625 RADGQSLHDSKQOTKILLITTKSDOMESPPLASSISDTNKNKSTGSLPSSGSTRGTSIK 664  
 DB 2261 DTVSYPQLH-LRQQRSDRDQMDPSQSVKPS--NEBGETKITLKLPMG----- 2306  
 QY 685 EKHKILIRLLDSS-SPVDLAKLTAETGKDLQESSSTAPGSEVITKQEPVAPKKENA 743  
 DB 2307 HRPSILHEHIGSSSRDRQKGRKSSKETCEKHSKSYLEPGOVYTGEBGNLPERADEV 2366  
 QY 744 LIRYLLDKDDTKDIGLPEITPKLERL-----DSKTDPAS-----NTRLIAMKTE 787  
 DB 2367 LTPGFLGRCNNVSSSEKIDKVLPLSGVPKGGSTQVEGSSKELQAPRKCSVAVTPLKME 2426  
 QY 788 KEKSEFPFGQPGSELNDEILIDLQNSQLPOLFPDTRGARA--GSVUKOAIINDLQ 845  
 DB 2427 GENOS-----KNTQ-----KESGGSFAHIESV----- 2449  
 QY 846 LTMENSPVTGVAOKTALRISQSTFNPRPQGLRLLPN-----QNLV-LDIT 892  
 DB 2450 -----CPAEVYSAR-----SPGAGPVGVSPNNTLSQDPQSNNTQNLPEDRN 2493  
 QY 893 LQSPFGAP-----FP-----PIRNSPY-----SVIPQPG 918  
 DB 2494 LMIIPDGKPKQEDGSFKRRYRARSARASNMFFGLTPLYGRVSGEEDIPYNSGTKRKG 2553  
 QY 919 MGNQGMIGNOGULGN-SIGMIGNSASRPTMSG--EMAPSSAVR--VTCAATTSAM 973  
 DB 2554 KRSAGGVGADGLDSTDEDLVYNYETRTVISGGEERLASHNLREEBQCOLPKISOL 2613  
 QY 974 RPYQGMIRNPASIPRPPSSQPGQROTLOSQVNI----- 1009  
 DB 2614 DGVDGTESTSVTATSRKSSQIPKRNKENGENTLKIDRPEADAGEKHYKSAVGHKNE 2673  
 QY 1010 -----GPSELEMMNGGPQYSQ-----QAPNPQT 1033  
 DB 2674 PKLDNCHSVSRVKAQGDGDSLEAQLSSLSRRVHTSPDSKNLIDRYNAELKSDSDNNN 2733  
 QY 1034 APMPESILPTIDQASFASQ--RQPESSPDLLCPHAPAESPSDEGALLDQYLALNR 1090  
 DB 2734 SDDCGNLPDIDMFYLNKTPSMQALGESP-----ESSSE-----LTLGEG 2776  
 QY 1091 DGLE-EIDRALGIPELVQSQAADPEQFSQDSINIMLEQ-----KAPV 1132  
 DB 2777 LGDLSNREKDIGLFYEVQSQDLPAPEVDSSVSSISAEBOFELPLRLPSDLSVLTTRSP 2836  
 QY 1133 FPOQYASQAQMAQS-----YSPMDP----- 1154  
 DB 2837 VPSQNPRLAVIIDSCEKRYITEKSVASSEGBPALSPVDVPADEGHMPTDFIQHMD 2896  
 QY 1155 -----NFTMGORPSTATLMDPRGLR---PTGLVQNPQNLRLQDLHRIDAQ 1201  
 DB 2897 ADHISPPCGSVQGHNSODLTRNSGTPGLQVPSVTPVQNO-----KVPSS 2946  
 QY 1202 NROPALNOISNVSNVNLTPRGVTPQPIVNAOMIAQORFILLNOHLROROMO-----Q 1255  
 DB 2947 TDSBGPQISNAA-----VTTTPHLKPATEKL-----VVAQNMOPLYVLOTLPNGVT 2995  
 QY 1256 QOVQOQRTLMRGGLNTP-----SMVAPSGMPATMS--NPRIPOA----- 1294  
 DB 2996 OKIQLTS-----PVSETPVMEIINTSVLGPMSGLTGLNPLSLPPSPSLPPASKGL 3049  
 QY 1295 -----NAQOFPPPN-----YGISQDPGFTGA-----TT 1320

```

Db 3050 LSVPHHQLHSFPAAQSSFPNISPSSGLLIGVQPPPDQLGSEANQRTDITTTVAT 3109
Oy 1321 POSPLASPRMA--HTOSPMQOOSANPAYQAPSDI-----NGMACGN 1360
Db 3110 PSSGKAKRPIISLHTRKKKALPSSAPSVIATSDVSNMTLINFTPSOLSNHPSLLDG 3169
Oy 1361 MGNSS-----MFSQSS-----PPHFGQOANTSMYSNNMINVSMATN---- 1397
Db 3170 LNPSSHRTVPNIRKSKSGIWMFEQAPLLPQSVGCTAATAGSSTISQDTISHLISGPVS 3229
Oy 1398 --TGSMSSMONTGOISMTSVTSVSTSG 1423
Db 3230 ALASGSSVLNVSMQTTAAPSTSTSVPG 3257

RESULT 40
TRX_DROME STANDARD; PRT; 3726 AA.
ID TRX_DROME STANDARD; PRT; 3726 AA.
AC P20659; Q27255; Q27327;
DT 01-FEB-1991 (Rel. 17, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Trithorax protein.
GN Trx.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
CX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90192757; PubMed=2107543;
RA Mazo A.M., Huang D.-H., Mozer B.A., David I.B.;
RT "The trithorax gene, a trans-acting regulator of the bithorax complex
in Drosophila, encodes a protein with zinc-binding domains.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:2112-2116(1990).
RN [2]
RP SEQUENCE FROM N.A., ALTERNATIVE SPLICING, AND CHARACTERIZATION.
RX MEDLINE=95009521; PubMed=7924966;
RA Sedkov Y., Tiliib S., Mizrokh L., Mazo A.;
RT "The bithorax complex is regulated by trithorax earlier during
Drosophila embryogenesis than is the Antennapedia complex, correlating
with a bithorax-like expression pattern of distinct early trithorax
transcripts.";
RL Development 120:1907-1917(1994).
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN=OREGON-R;
RA MEDLINE=96100387; PubMed=8555104;
RA Tiliib S., Sedkov Y., Mizrokh L., Mazo A.;
RT "Conservation of structure and expression of the trithorax gene
between Drosophila virilis and Drosophila melanogaster.";
RL Mech. Dev. 53:113-122(1995).
RN [4]
RP CHARACTERIZATION.
RX MEDLINE=95047388; PubMed=7958911;
RA Kuzin B., Tiliib S., Sedkov Y., Mizrokh L., Mazo A.;
RT "The Drosophila trithorax gene encodes a chromosomal protein and
directly regulates the region-specific homeotic gene fork head.";
RL Genes Dev. 8:2478-2490(1994).
CC -1- FUNCTION: FUNCTIONS IN SEGMENT DETERMINATION THROUGH INTERACTION
WITH GENES OF BITHORAX (BX-C) AND ANTENNAPEDIA (ANT-X) COMPLEXES.
IT CAN BEHAVE AS AN ACTIVATOR OF BX-C.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- MISCELLANEOUS: THIS PROTEIN HAS BEEN EXPERIMENTALLY SHOWN TO BIND
ZINC.
CC -1- SIMILARITY: BELONGS TO THE TRANSCRIPTION FACTOR TRITHORAX FAMILY.
CC -1- SIMILARITY: CONTAINS 5 PHD-TYPE ZINC FINGERS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; M31617; AAA29025.1; -
DR EMBL; Z50152; CAA90514.1; -
DR EMBL; Z50152; CAA90513.1; -
DR EMBL; Z31725; CAA83516.1; -
DR EMBL; Z31725; CAA83515.1; -
DR PIR; A35085; A35085.
DR HSSP; P19793; 2NLL.
DR TRNSPAC; T00850; -
DR FlyBase; FBgn0003862; trx.
DR InterPro; IPR003889; Fyricl_C.
DR InterPro; IPR003888; Fyricl_N.
DR InterPro; IPR001965; PHD.
DR InterPro; IPR003616; PostSET.
DR InterPro; IPR001214; SET.
DR InterPro; IPR001841; Znf_ring.
DR Pfam; PF00628; PHD; 2.
DR Pfam; PF00856; SET; 1.
DR SMART; SM00542; FYRC; 1.
DR SMART; SM00541; FYRN; 1.
DR SMART; SM00249; PHD; 4.
DR SMART; SM00308; PostSET; 1.
DR SMART; SM00184; RING; 3.
DR SMART; SM00317; SET; 1.
DR PROSITE; PS50280; SET; 1.
KW Transcription regulation; Zinc-finger; Metal-binding; DNA-binding;
KW Nuclear protein; Developmental protein; Activator;
KM Alternative splicing.
FT ZN_FING 1266 1347 PHD-TYPE 1.
FT ZN_FING 1348 1393 PHD-TYPE 2.
FT ZN_FING 1421 1482 PHD-TYPE 3.
FT ZN_FING 1734 1793 PHD-TYPE 4 (ATYPICAL).
FT ZN_FING 1794 1844 PHD-TYPE 5 (ATYPICAL).
FT DOMAIN 3599 3708 SET.
FT DOMAIN 512 516 POLY-SER.
FT DOMAIN 565 570 POLY-ASP.
FT DOMAIN 661 664 POLY-SER.
FT DOMAIN 905 910 POLY-SER.
FT DOMAIN 1576 1582 POLY-GLN.
FT DOMAIN 2298 3027 GLN-RICH.
FT DOMAIN 3032 3040 POLY-SER.
FT DOMAIN 3181 3184 POLY-GLN.
FT DOMAIN 3220 3225 POLY-GLU.
FT VARSPLIC 1 368 MISSING (IN SHORT ISOFORM).
FT CONFLICT 2025 P -> PWLTSPKFLGLSTHGGLLMLLGVVRLKGG
(IN REF. 1).
FT CONFLICT 2341 R -> S (IN REF. 1).
FT CONFLICT 2392 G -> S (IN REF. 1).
SQ SEQUENCE 3726 AA; 400575 MW; D2756E50763DLCF5 CRC64;

Query Match 3.2%; Score 241.5; DB 1; Length 3726;
Best Local Similarity 18.7%; Pred. No. 0.0027;
Matches 294; Conservative 204; Mismatches 596; Indels 477; Gaps 69;

Oy 108 GIYDDALGPMLEALDGFPPVNI--GNVPVSENVOTLRKNOEELM--MKSYSI 162
Db 1890 GSEVRQLGATVPRESDSYEAVPIPLFCISRLYSSKEPKIVETVFTTIONSSTTLTA 1949
Oy 163 LHVQ-----DHT-----EFYKNLLPSIYVGGSGMSEPPRR 193
Db 1950 LDVGRNYTVDHNPNSKEVQGMQAIARHNSLASSELEN-----GCTWSEFFNP 2002
Oy 194 NSHTNCRMLVKPLPDSEEGHNOE-----AHQKETMOCFAV-----S 233
Db 2003 NS-----CVPDPONTFEEEPQOQADLLPPELKDAIFEDLPHELDGISMLDIFYDXT 2055

```

```

QY 234 QPKSKEEEDLOSLCICVARVPKEREVLPSESEFTTRDLOG---KITSIDTSMR 289
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2056 DLPAISEOSK-----GTQAMTSNOAQONQOAGANSVSIICEDETRNSN 2100
QY 290 AAKKPPG-----EDLVRCIOKFNHQHEGESYAKRHHHEVLRQCLASQIYREFLSD 343
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2101 TSIJGNGPASPVEDDALSAARNSOYOMLKTAMPKLDGNSAATAIKRRLSK-LWAE 2159
QY 344 GTLVAQOTSKLIRSOTTNEPOLIVISLHMLREQONVCVMPNDLTGTMRKPLNFISNS- 402
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2160 GVELT-----LSSQORNNKEMA-TYAGVSRQISISETSEVGAITSGS---VRKSF 2207
QY 403 --PAHQALCSGNPGQDMLTSSNIFPINGKEQMGMPMRGSGGKN--HVSGM----- 453
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2208 TWSAAKRYFEKSEGREEAAMRI-MQMDGYDDSI-TEFRITISGGINSTAQSGQVXCDR 2265
QY 454 -QATTPGGSNY-ALKMSPGSSPGMNPQGTSLSPRHMSPEVACSPHPSQFPAG 511
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2266 CQCTY---RNYDAFQRLPS-CSFTMSSNETESDVSGO----- 2239
QY 512 SLHSPVGCSTGNSHSYTNSSLNAL--QALSEGHVSLGSSLASPDLKMGNLQNSPVM 569
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2300 -----GMTNNATQISAESLNELOKOLLANAGLNYLOSATS----- 2335
QY 570 NPPPLSMGSLDKCFGLYG-----EPSEGTGAESSCHGEQKEIND----- 614
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2336 -FPOVORLGSIGQ---FGLQGLQOLQLOPQSLNGFELSQPNPQANTDLOIYANSIQ 2391
QY 615 ---PNL-----PRVSSERADQSR-----HDSKGQTKLQULTTKSDQMPSPCLASS 660
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2392 GLAANLGGGTTLAOPYT---APAQPOLIAVSTNPDSQOPIQIPQTMQATTPETARYQT 2448
QY 661 LSDTNKSDSTSLPGSGS---THGTSLEKHKILHRLQDSS--SPVDLAKTLAEATGK 713
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2449 LQATNTDKKLMLELTAGKRLKLTATKAAQQAANKQRLKSGHGVKRIQ-AKLOPHRQH 2507
QY 714 DLSQESSSTAPGSEVTIKQEPVSPKKENALLRYLKDQDTK---DIGLEITPKLERL 769
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2508 QOQOQ-----TQVOQPI-TVMQONLQPOLLFQSSQTQAPQIILPQAP----- 2551
QY 770 DSKTDPASNTKILAMKTEKEMSEPDQSGELDNLEIILDLQNSQLPQ-----LF 822
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2552 -----QNTISFTGD--GSQOPLQYISIPAGYKPKPOPTATPTF 2591
QY 823 PDTRPGAPAGSVDKOALINILMOILTANSEFTVPYGAQKTALRISQSTFNPNRPOLGRL 882
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2592 LTTAPGAGATYLTQDASGNLVITTPNSGLQMLTAQSL-----QAQPVIGTLI 2641
QY 883 PNQNLPLDITLOSPTGAGPPPIRNSP-----XSVIPQGMG 921
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2642 QPQ-----TLOLGGADGNQPSNOQPLILGTGGSGGLEFATTSPOVILATQPMYXG 2695
QY 922 NOGIGN-----QGNLGNSTGMIGNSASRPMPSCGEMAPQSSAVRYTCAATTSAMNR-- 974
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2696 LETIVONTVWSSQOQFVSTAMPGLSONASFSATTTQVFOASKIEPIYDLPAQIVVLNNTG 2755
QY 975 -PVQGMIRNPASISIPRPSOPQROTLOSOVANIGPSELEMMNGPQYSSQOQAPPNOT 1033
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2756 DASSAGTEFLN-AASVLTQOQODDTTQILQN--ANFQFQSVPTSSGASTMDYTSPIVMYV 2812
QY 1034 APMPESLPLIDQAFASQNPQFGSSPDLLCPHPAAESPDSGALLDOLYLALRNFGL 1093
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2813 AKIP---PVTOIKRTNAQAKAAGISGVGVPPQ----- 2843
QY 1094 EETIDRALGIEIYSSQAVPEQFSSQDSDNIMLEOKAPVFPQYASQQAQSGSYSPMD 1153
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2844 QYVANKVLPSTIVTQSOV-----QVXNSNLQ-----SOYKGAASGTGT----- 2883
QY 1154 PNFTMGQRPSTATLRMQPRG-----LRPTGLVONQPNQRLQLOHRLQA 1199
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2884 ----TGAPPSIASKPLQKTKTNNIRPIHKLEVVKPVKVKPTPKVQONHSL-----LQO 2932
QY 1200 QONRQPLANO-----ISNVSNVNLIRPGVPTQAPINAQMLAOROREILNQH-LRQROM 1252

```

```

Db 2933 QOQOQPOLQOQIIPAVVNVNOVPKYTI-----SQORIPAOTQOQLOQAQOMIHIPQOQCP 2985
QY 1253 HQOQOVO-----ORTLMRGQGLNM-----TPSNVAPSGMP 1283
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2986 LQOQOVOVQSPMLITLAEAPVVOQOFVMEPQALQEOELANRYOHTSSSSSSSSCQSLP 3045
QY 1284 ATMSNP---RIPQANAQOFPPPNYGI---SQQPD-----GFTGATTPQSPMLSPRMA 1331
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3046 TNVNVPMQQAAPSTSSSTTRPTNRVLPQOQROQEPAPLSNECPVSSPFPKPEQPIIH 3105
QY 1332 HTQSPMAQSOA-----NPAIQAPSDINGMAQNGGNSM--FSQOSPPHFQOQANTS 1383
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3106 QMTSASVSKCYAQKSTLPSPEYAELEKVSLSLEIYDPVTMDAILREQP-----VTQSI 3159
QY 1384 YSNMNNINVMATNTGMSMSMNOQISMTSVTSVSTGLSSMGPROVNDPALRGCLNF 1443
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3160 YTEGL-----YEKNSPEESTBOLL-----LQOQORQLNQOLVNNGYLL 3199
QY 1444 PN---QLPGMD 1451
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3200 DKHTFOVEPMD 3210

```

Search completed: September 7, 2002, 10:44:39  
 Job time: 370 sec

This Page Is Inserted by IFW Operations  
and is not a part of the Official Record

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

**IMAGES ARE BEST AVAILABLE COPY.**

As rescanning documents *will not* correct images,  
please do not report the images to the  
Image Problem Mailbox.

**THIS PAGE BLANK (USPTO)**